PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] : C12N 15/12, C07K 14/47, C12Q 1/68, C07K 16/18, C12N 15/11	A2	(11) International Publication Number: **WO 99/13075** (43) International Publication Date: 18 March 1999 (18.03.99)

(21) International Application Number: PCT/US98/18638

(22) International Filing Date: 8 September 1998 (08.09.98)

(30) Priority Data:
60/058,180 8 September 1997 (08.09.97) US
60/059,725 22 September 1997 (22.09.97) US

(71) Applicant: PRINCETON UNIVERSITY [US/US]; New South Building, 5th floor, P.O. Box 36, Princeton, NJ 08544-0036 (US).

(72) Inventors: ZHU, Hua; 7-U Hibben Apartment, Faculty Road, Princeton, NJ 08540 (US). CONG, Jiang-Ping; 7-U Hibben Apartment, Faculty Road, Princeton, NJ 08540 (US). SCHENK, Thomas; 87 McCosh Circle, Princeton, NJ 08540 (US).

(74) Agents: JACKSON, David, A. et al.; Klauber & Jackson, 411 Hackensack Avenue, Hackensack, NJ 07601 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: HUMAN GENES REGULATED BY HUMAN CYTOMEGALOVIRUS AND INTERFERON

(57) Abstract

This invention describes 23 genes related to HCMV infection. We demonstrate that 19 out of the 23 genes are induced by HCMV infection and that 6 of these genes were previously unidentified in public sequence data bases. We also show for the first time that 7 out of the 19 genes are also interferon-inducible. We demonstrate that 4 out of the 23 genes are repressed by HCMV infection. Two of these are known genes and the cDNA sequence that we have determined for the other two are not present in public data bases. The invention relates to using these genes as markers in assays screening for compounds that reverse the expression pattern of said genes following challenge with either cytomegalovirus or interferon. The invention further relates to anti-viral pharmaceutical compositions encompassing recombinant proteins, antibodies, antisense technology and gene therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

HUMAN GENES REGULATED BY HUMAN CYTOMEGALOVIRUS AND INTERFERON

TECHNICAL FIELD OF THE INVENTION

The present invention relates to the identification of genes in which their expression is either induced or repressed upon either cytomegalovirus infection or interferon treatment. The invention also relates to using these genes as markers in assays screening for compounds that reverse the expression pattern of said genes following challenge with either cytomegalovirus or interferon. The invention further relates to anti-viral pharmaceutical compositions enncompassing recombinant proteins, antibodies, antisense technology, and gene therapy.

BACKGROUND OF THE INVENTION

Human cytomegalovirus (HCMV) is a wide-spread human pathogen that causes birth defects and can be life-threatening to people whose immune system is compromised (AIDS and transplant patients). HCMV can alter gene expression through multiple pathways. For example, the virion gB and gH glycoproteins induce cellular transcription factors when they interact with their cell surface targets (1). Virion proteins, such as pp71 (2-4), can activate transcription (5); and viral proteins synthesized after infection, such as IE1 and IE2, regulate expression from a variety of promoters (6-10). Further, HCMV infection has been shown to perturb cell cycle progression (11-14), which leads to changes in gene expression.

Viral factors, induced cellular factors and changes in cell cycle progression have the potential to exert profound effects on gene expression, but relatively few cellular genes have been identified whose activity changes after HCMV infection (15). A more global understanding of HCMV-induced changes in cellular gene expression should help us to better understand how the virus interacts with its host cell during

the replication process, and might direct us to new targets for therapeutic intervention in HCMV disease.

SUMMARY OF THE INVENTION

In accordance with the present invention, certain novel cDNA sequences have been identified that originate from mRNAs that are expressed in response to HCMV infection. Therefore, the genes that encode these mRNAs are termed HCMV *i*nducible *g*enes (*cig*). Interestingly, and as set forth herein, these genes were also found to be inducible by interferon-α.

Accordingly, 19 genes that are induced upon HCMV infection of human cells and 4 genes that are repressed by HCMV infection of human cells have been identified. Further, the present invention reveals that the genes which are induced by HCMV infection are also induced by interferon-α. Finally, the 19 genes that are induced by HCMV and interferon-α include 6 genes that have not been reported previously.

Also in accordance with the present invention, certain novel cDNA sequences have been identified that originate from mRNAs that are repressed in response to HCMV infection. Therefore, the genes that encode these mRNAs are termed HCMV *r*epressable *g*enes (*crg*).

In one embodiment of the invention, the *cigs* can be used as markers for use in a screening assay to identify compounds that prevent the expression of any of these genes. Likewise, the *crgs* can also be used as markers for use in a screening assay to identify compounds that relieve the repression of these genes.

In a further embodiment, the screening assays also extend to use of antibodies against the proteins encoded by the above-mentioned cDNAs in an ELISA-type assay.

In a yet a further embodiment, the screening assays can also be used to follow the efficacy of various treatment regimens in patients, thus leading to more effective treatment.

The present invention also extends to therapeutic applications utilizing the nucleotide sequences derived from the *cigs* and *crgs* in antisense therapeutics and gene therapy.

In a further aspect, the encoded proteins that can be infered from the cDNA sequences of the *cigs* and *crgs* can also be used in therapeutic applications. The fact that the *cigs* are also induced by interferon, combined with the fact that interferons are used in anti-viral therapy, gives strength to the notion that the proteins have potential as generic anti-viral compounds.

In yet a further aspect, one or more of the encoded proteins from the *cigs* may be responsible for the toxicity of interferon. Therefore, the newly discovered gene products have utility as targets for screens to discover compounds that could block this toxicity, thus leading to drugs that could greatly enhance the efficacy of interferon treatment by allowing the use of higher doses of interferon.

In a particular embodiment, the present invention relates to all members of the herein disclosed family of *cigs* and *crgs*.

The present invention also relates to a recombinant DNA molecule or cloned gene, or a degenerate variant thereof, which encodes any *cig* or *crg* gene product; preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the *cig* or *crg* gene product has a nucleotide sequence or is complementary to a DNA sequence contained in any of the *cigs* or *crgs* identified in the Sequence Listing as SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39.

The human and murine DNA sequences of the *cigs* and *crgs* of the present invention or portions thereof, may be prepared as probes to screen for complementary sequences and genomic clones in the same or alternate species. The present invention extends to probes so prepared that may be provided for screening cDNA and genomic libraries for the *cigs* and *crgs*. For example, the probes may be prepared with a variety of known vectors, such as the phage λ vector. The present invention also includes the preparation of plasmids including such vectors, and the use of the DNA sequences to construct vectors expressing antisense RNA or ribozymes which would attack the mRNAs of any or all of the DNA sequences set forth in the Sequence Listing (SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39). Correspondingly, the preparation of antisense RNA and ribozymes are included herein.

The present invention also includes *cig* or *crg* gene products (*i.e.* proteins) having the activities noted herein, and that contain amino acid sequences set forth in the Sequence Listing and selected from SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 27, 29, 31, 33, 35, and 37.

In a further embodiment of the invention, the full DNA sequence of the recombinant DNA molecule or cloned gene so determined may be operatively linked to an expression control sequence which may be introduced into an appropriate host. The invention accordingly extends to unicellular hosts transformed with the cloned gene or recombinant DNA molecule comprising a DNA sequence encoding any one of the present *cigs* or *crgs*, and more particularly, the complete DNA sequence determined from the sequences set forth above and in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39.

According to other preferred features of certain preferred embodiments of the present invention, a recombinant expression system is provided to produce biologically active animal or human *cig* or *crg* gene products.

The present invention naturally contemplates several means for preparation of the *cig* or *crg* genes and gene products, including as illustrated herein known recombinant techniques, and the invention is accordingly intended to cover such synthetic preparations within its scope. The isolation of the cDNA and amino acid sequences disclosed herein facilitates the reproduction of the *cigs* and *crgs* by such recombinant techniques, and accordingly, the invention extends to expression vectors prepared from the disclosed DNA sequences for expression in host systems by recombinant DNA techniques, and to the resulting transformed hosts.

The invention includes an assay system for screening of potential drugs effective to modulate *cig* or *crg* expression levels of target mammalian. In one instancé, the test drug could be administered to a cellular sample, prior to or after HCMV-infection or interferon treatment, to determine its effect upon the *cig* or *crg* expression level to any chemical sample (including DNA), or to the test drug, by comparison with a control.

The assay system could be adapted to identify drugs or other entities that are capable of reducing the toxicity of interferon treatment by antagonizing one or more of the *cigs*. Such assay would be useful in the development of drugs that would allow for higher dosage interferon treatments without the concomitant toxicity normally associated with administering high levels of interferon.

In yet a further embodiment, the invention contemplates antagonists of the activity of a *cig* gene product. In particular, an agent or molecule that inhibits any *cig* gene product and, in turn, has antiviral activity in general and anti-HCMV activity in particular.

In still yet a further embodiment, the invention contemplates the use of a *crg* gene product as a therapeutic to treat HCMV infection. As infection with HCMV reduces the level of these gene products, it follows that replacement of this gene

product, either through gene therapy or *via* direct administration of the gene product, has potential to alleviate HCMV infection and/or its associated symptoms.

The present invention extends to the development of antibodies against the *cig* or *crg* gene products, including naturally raised and recombinantly prepared antibodies. For example, the antibodies could be used to screen expression libraries to obtain the gene or genes that encode the *cig* or *crg* gene products. Such antibodies could include both polyclonal and monoclonal antibodies prepared by known genetic techniques, as well as bi-specific (chimeric) antibodies, and antibodies including other functionalities suiting them for additional diagnostic use conjunctive with their capability of modulating activities associated with the *cig* or *crg* gene products.

Thus, *cig* or *crg* gene products, their analogs and/or analogs, and any antagonists or antibodies that may be raised thereto, are capable of use in connection with various diagnostic techniques, including immunoassays, such as a radioimmunoassay, using for example, an antibody to the *cig* or *crg* gene products that has been labeled by either radioactive addition, or radioiodination.

In an immunoassay, a control quantity of the antagonists or antibodies thereto, or the like may be prepared and labeled with an enzyme, a specific binding partner and/or a radioactive element, and may then be introduced into a cellular sample. After the labeled material or its binding partner(s) has had an opportunity to react with sites within the sample, the resulting mass may be examined by known techniques, which may vary with the nature of the label attached.

In the instance where a radioactive label, such as the isotopes ^{3}H, ^{14}C, ^{32}P, ^{33}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re are used, known currently available counting procedures may be utilized. In the instance where the label is an enzyme, detection may be accomplished by any of the presently utilized

colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques known in the art.

The present invention includes an assay system which may be prepared in the form of a test kit for the quantitative analysis of the extent of the presence of the *cig* or *crg* gene products (either mRNA or protein), or to identify drugs or other agents that may mimic or block their activity. The system or test kit may comprise a labeled component prepared by one of the radioactive and/or enzymatic techniques discussed herein, coupling a label to the *cig* or *crg* gene products, their agonists and/or antagonists, and one or more additional immunochemical reagents, at least one of which is a free or immobilized ligand, capable either of binding with the labeled component, its binding partner, one of the components to be determined or their binding partner(s).

In a further embodiment, the present invention relates to certain therapeutic methods which would be based upon either modulating expression levels *cig*s and/or *crg*s or antagonizing the activity of any of the *cig* gene products, their subunits, or active fragments thereof, or upon agents or other drugs determined to possess the same activity. A first therapeutic method is associated with the prevention of the manifestations of conditions causally related to or following from HCMV infection, and comprises administering an agent capable of modulating the production and/or activity of any of the *cig* or *crg* gene products, either individually or in mixture with each other in an amount effective to prevent the development of those conditions in the host. For example, drugs or other binding partners to the *cig* or *crg* gene products may be administered to inhibit or potentiate their activity, as it relates to HCMV or other viral infection.

Accordingly, it is a principal object of the present invention to provide *cig* or *crg* gene products in purified form that have utility in treating, or identifying drugs (compounds) to treat, HCMV or other viral infection.

It is a further object of the present invention to provide antibodies to the *cig* or *crg* gene products, and methods for their preparation, including recombinant means.

It is a further object of the present invention to provide a method for detecting the presence of the *cig* or *crg* mRNA or protein gene products in mammals in which invasive, spontaneous, or idiopathic pathological states are suspected to be present.

It is a further object of the present invention to provide a method and associated assay system for screening substances such as drugs, agents and the like, potentially effective in either mimicking the activity or combating the adverse effects of the *cig* or *crg* gene products in mammals.

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the *cig* or *crg* gene products, so as to alter the adverse consequences of such presence or activity, or where beneficial, to enhance such activity.

It is a still further object of the present invention to provide a method for the treatment of mammals to control the amount or activity of the *cig* or *crg* gene products, so as to treat or avert the adverse consequences of invasive, spontaneous or idiopathic pathological states.

It is a still further object of the present invention to provide pharmaceutical compositions for use in therapeutic methods which comprise or are based upon the *cig* or *crg* gene products, their binding partner(s), or upon agents or drugs that control the production, or that mimic or antagonize the activities of the *cig* or *crg* gene products.

Other objects and advantages will become apparent to those skilled in the art from a review of the ensuing description which proceeds with reference to the following illustrative drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1. Characterization of UV-inactivated HCMV (UV HCMV). (A) Western blot showing that UV irradiation of the virus blocks expression of the HCMV IE1 and IE2 RNAs, but has no effect on the delivery of a virion protein to cells. HF cells were mock infected or infected, and extracts were prepared 8 or 21 h later. Lanes 1-6 were reacted with an antibody (MAb810) that binds to two immediate early proteins (IE1 and IE2), while lanes 7-9 were reacted with an antibody to a virion constituent, pp65. The molecular weights of marker proteins are indicated to the left of the panels. (B) Northern blot showing that IE1 RNA is detected at 8 h after infection of HF cells with HCMV but not after infection with UV HCMV. (C) Immunofluorescent localization of pp65 and IE1/2 within infected cells. HF cells were infected with for 2 or 8, reacted with antibody to pp65 or IE1/IE2 followed by a fluorescein-labeled secondary antibody and counterstained with ethidium homodimer-1.

FIGURE 2. Differential expression of RNAs in HF cells assayed by Northern blot. (A) RNA was prepared from mock-infected (M), HCMV-infected cells (C), or UV HCMV-infected cells and assayed using cloned cDNA segments. The different clones (*cigs*) are identified above the panels. (B) RNA was prepared from mock-infected (M), HCMV-infected (C) or HCMV-infected cells that were treated with cycloheximide (CX) and assayed as in panel A. (C) RNA was prepared from mock-infected cells (M), HCMV-infected cells (C) or cells treated with interferon-α (I) and assayed using probes corresponding to the cigz or the previously characterized, interferon-inducible mxA gene. (D) RNA was prepared as in panel A and assayed using probes corresponding to known interferon-inducible genes (mxA, isg15K, IFN-b) or control genes that are not induced by interferon (p53, p21, cPLA2, actin).

FIGURE 3. The HCMV particle mediates the induction of differentially expressed HF RNAs. (A) Requirements for induction monitored by Northern blot assay. The relative amounts of three cellular RNAs (*cig*1, *cig*6 and *cig*49) were monitored in mock-infected cells (mock), HCMV strain AD169-infected cells (HCMV), medium from a virus stock from which HCMV particles were removed by filtration (inf. med.), mock-infected cells treated with 100 mg/ml cycloheximide (CX), HCMV strain AD169-infected cells treated with 100 mg/ml cycloheximide (CX+HCMV), cells infected with purified HCMV strain AD169 particles (virions), cells infected with purified non-infectious enveloped particles from strain AD169 (NIEPs), adenovirus-infected cells (Ad *dl*309), herpes simplex type 1-infected cells (HSV-1), HCMV strain Towne-infected cells, HCMV strain Toledo-infected cells,interferon-α-treated cells (IFN-α), and interferon added to medium and passed through the filter to exclude virus (fil. IFN-α). (B) Northern blot assay demonstrating that antibody which neutralizes HCMV (antibody C) blocks the induction of *cig* RNA accumulation, while antibodies that neutralize interferon-α or β (antibody Iα, Iβ) block the induction of *cig* RNAs by interferon-α or β (inducer Iα, Iβ) but have no effect on the induction of *cig* RNAs by HCMV (C). RNA prepared from mock-infected control cells is designated M. The cellular cytosolic phospholipase A2 RNA that is not modulated by infection or interferon treatment is assayed at the bottom of the figure as a loading control (control, cPLA2).

FIGURE 4. Requirements for the induction of *cig* RNA accumulation. (A) An intact HCMV particle is required. Purified HCMV particles were treated with a mixture of TritonX100 and DOC (T/C) and separated by centrifugation into supernatant (S) and pellet (P) fractions. Northern blot assays show the effect of detergent treatment on the induction of two *cig* RNAs (*cig*1 and *cig*49) by virus particles (HCMV) or interferon-α (IFN-α). (B) The induction of *cig* RNAs does not invovle the release and subsequent action of mediators stores within infected HF cells. At 8 h after treatment, RNA was prepared from mock-infected cells (lane 1), HCMV-infected cells (lane 2), or a 9:1 mixture of mock and infected cells (lanes 3

and 4). The two mixed cultures differed in the time after infection when the cells were mixed. In mixture 3, cells were mixed at 1 h after infection; in mixture 4, RNAs were prepared and mixed from 8 h mock- and HCMV-infected cells. RNAs were analyzed by Northern blot using cellular (*cig*1, *cig*6, *cig*49, cPLA2) and viral (IE1) probes.

FIGURE 5. Kinetic analysis of *cig* RNA accumulation. HF cells were either mock-infected (M) or treated with the inducers identified to the right of each blot (HCMV, HSV-1, IFN-α), RNA was prepared at various times after treatment (indicated above lanes), and analyzed by Northern blot using the probes indicated to the left of each blot (*cig*1, *cig*49, HCMV IE1, HSV-1 icp47).

DETAILED DESCRIPTION

As described in detail *infra*, differential display analysis was employed to identify mRNAs that accumulate to enhanced levels in human cytomegalovirus-infected as compared to mock-infected cells. RNAs were compared at 8 hours after infection of primary human fibroblasts. Fifty-seven partial cDNA clones were isolated, representing about 26 differentially expressed mRNAs. Eleven of the mRNAs were virus-coded and 15 were of cellular origin. Six of the partial cDNA sequences have not been reported previously. All of the cellular mRNAs identified in the screen are induced by interferon-α and β. The induction in virus-infected cells, however, does not involve the action of interferon or other small signalling molecules.
Neutralizing antibodies that block virus infection also block the induction. These RNAs accumulate after infection with virus that has been inactivated by treatment with UV light, indicating that the inducer is present in virions. From the above, it is concluded that human cytomegalovirus induces interferon-responsive mRNAs.

In its broadest aspect, the invention describes 23 genes related to HCMV infection. These genes are described in the EXAMPLES. We show for the first time that 19 genes are induced by HCMV infection (see Table 1 in EXAMPLE 4); we identify 6/19 genes for the first time (these genes are listed as "new" in Table 1), *i.e.*, the partial cDNA sequences that we have derived are not found in public sequence data bases; 12/19 genes were previously shown to be induced by interferon, and we show for the first time that 7/19 genes are induced by interferon (the 6 genes listed as "new" in Table 1 as well as KIAA0062).

Since these genes are expressed at high level in HCMV-infected cells, it is possible that they are needed for successful replication and spread by the virus. Therefore, the genes have utility as targets for the development of screens to identify drugs that inhibit their expression or action. Inhibition of the normal activity of these HCMV-induced cellular gene products might inhibit HCMV replication and spread. It may also be possible to identify the viral gene product that causes the enhanced expression of these genes and discover a drug that blocks its function, thereby preventing accumulation of these cellular genes.

The 7 genes that are shown to be induced by interferon-α for the first time have additional utility. This is probably the most important aspect of the invention since interferon-related activities are not limited to the control of HCMV. Interferons alpha and beta exhibit many different functions, including: (1) the induction of an antiviral state; (2) inhibition of cell growth; (3) induction of class I MHC antigens; and (4) activation of macrophages, natural killer cells and cytotoxic T lymphocytes. Interferons can block the replication and spread of many different viruses, the growth of nonviral pathogens and the growth of certain cancer cells. Interferon functions by initiating a signaling cascade that results in the expression of interferon-responsive gene products that then mediate interferon actions, such as antagonizing the growth of a virus (given this function of interferon, it is strange that HCMV induces interferon-response genes). The 7 newly identified gene

products could exhibit subsets of the activities ascribed to interferons alpha and beta. Therefore, they have potential as therapeutic proteins. The utility of interferons as therapeutic agents is limited because they are toxic. Possibly one or more of these newly discovered interferon-response genes produces a product that is responsible for the toxicity (or a significant portion of the toxicity). If so, the newly discovered gene products have utility as targets for screens to discover drugs that could block aspects of their activity that leads to toxicity. Such drugs could greatly enhance the utility of interferons as therapeutics by reducing their toxicity and permitting higher doses.

We show for the first time 4 genes that are repressed by HCMV infection. Two of these are known genes and the cDNA sequence that we have determined for the other two are not present in public data bases. If their repression is important for HCMV replication and spread, then the delivery of these products as proteins or perhaps within an expression vector could interfere with HCMV replication and spread. It might also be possible to identify the viral gene product that is responsible for their repression and discover a drug that blocks its function.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook et al, "Molecular Cloning: A Laboratory Manual" (1989); "Current Protocols in Molecular Biology" Volumes I-III [Ausubel, R. M., ed. (1994)]; "Cell Biology: A Laboratory Handbook" Volumes I-III [J. E. Celis, ed. (1994))]; "Current Protocols in Immunology" Volumes I-III [Coligan, J. E., ed. (1994)]; "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" [B.D. Hames & S.J. Higgins eds. (1985)]; "Transcription And Translation" [B.D. Hames & S.J. Higgins, eds. (1984)]; "Animal Cell Culture" [R.I. Freshney, ed. (1986)]; "Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

Therefore, if appearing herein, the following terms shall have the definitions set out below.

The term "*cig*" or "*cigs*" refers to HCMV-inducible genes.

The term "*crg*" or "*crgs*" refers to HCMV-repressable genes.

The nucleotide sequences of the cDNA molecules associated with the *cigs* and *crgs* is presented in the Sequence Listing (SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39).

The term "product" in "*cig* or *crg* gene product", and varients thereof, can refer to either protein or mRNA.

The term "*cig* or *crg* gene product(s)," and any variants not specifically listed, as used throughout the present application and claims can refer to proteinaceous material including single or multiple proteins, and extends to those proteins having the amino acid sequence data described herein and presented in the Sequence Listing (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 27, 29, 31, 33, 35, and 37), and the profile of activities set forth herein and in the Claims. Accordingly, proteins displaying substantially equivalent or altered activity are likewise contemplated. These modifications may be deliberate, for example, such as modifications obtained through site-directed mutagenesis, or may be accidental, such as those obtained through mutations in hosts that are producers of the complex or its named subunits..

The amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired fuctional property of immunoglobulin-binding is retained by the polypeptide. NH_2 refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group

present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J. Biol. Chem.*, **243**:3552-59 (1969), abbreviations for amino acid residues are shown in the following Table of Correspondence:

TABLE OF CORRESPONDENCE

SYMBOL		AMINO ACID
1-Letter	3-Letter	
Y	Tyr	tyrosine
G	Gly	glycine
F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

It should be noted that all amino-acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-

terminus to carboxy-terminus. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino-acid residues. The above Table is presented to correlate the three-letter and one-letter notations which may appear alternately herein.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in its either single stranded form, or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA).

An "origin of replication" refers to those DNA sequences that participate in DNA synthesis.

A DNA "coding sequence" is a double-stranded DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control of

appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences are DNA regulatory sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for the expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

An "expression control sequence" is a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then translated into the protein encoded by the coding sequence.

A "signal sequence" can be included before the coding sequence. This sequence encodes a signal peptide, N-terminal to the polypeptide, that communicates to the host cell to direct the polypeptide to the cell surface or secrete the polypeptide into the media, and this signal peptide is clipped off by the host cell before the protein leaves the cell. Signal sequences can be found associated with a variety of proteins native to prokaryotes and eukaryotes.

The term "oligonucleotide," as used herein in referring to the probe of the present invention, is defined as a molecule comprised of two or more ribonucleotides, preferably more than three. Its exact size will depend upon many factors which, in turn, depend upon the ultimate function and use of the oligonucleotide.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product, which is complementary to a nucleic acid strand, is induced, i.e., in the presence of nucleotides and an inducing agent such as a DNA polymerase and at a suitable temperature and pH. The primer may be either single-stranded or double-stranded and must be sufficiently long to prime the synthesis of the desired extension product in the presence of the inducing agent. The exact length of the primer will depend upon many factors, including temperature, source of primer and use of the method. For example, for diagnostic applications, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 or more nucleotides, although it may contain fewer nucleotides.

The primers herein are selected to be "substantially" complementary to different strands of a particular target DNA sequence. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For

example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to hybridize therewith and thereby form the template for the synthesis of the extension product.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

Two DNA sequences are "substantially homologous" when at least about 75% (preferably at least about 80%, and most preferably at least about 90 or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular

system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Maniatis et al., *supra*; DNA Cloning, Vols. I & II, *supra*; Nucleic Acid Hybridization, *supra*.

It should be appreciated that also within the scope of the present invention are DNA sequences encoding *cig* and *crg* gene products which code for proteins having the same amino acid sequence as SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 27, 29, 31, 33, 35, and 37, but which are degenerate to SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39. By "degenerate to" is meant that a different three-letter codon is used to specify a particular amino acid. It is well known in the art that the following codons can be used interchangeably to code for each specific amino acid:

Phenylalanine (Phe or F)	UUU or UUC
Leucine (Leu or L)	UUA or UUG or CUU or CUC or CUA or CUG
Isoleucine (Ile or I)	AUU or AUC or AUA
Methionine (Met or M)	AUG
Valine (Val or V)	GUU or GUC of GUA or GUG
Serine (Ser or S)	UCU or UCC or UCA or UCG or AGU or AGC
Proline (Pro or P)	CCU or CCC or CCA or CCG
Threonine (Thr or T)	ACU or ACC or ACA or ACG
Alanine (Ala or A)	GCU or GCG or GCA or GCG
Tyrosine (Tyr or Y)	UAU or UAC
Histidine (His or H)	CAU or CAC
Glutamine (Gln or Q)	CAA or CAG
Asparagine (Asn or N)	AAU or AAC
Lysine (Lys or K)	AAA or AAG
Aspartic Acid (Asp or D)	GAU or GAC
Glutamic Acid (Glu or E)	GAA or GAG
Cysteine (Cys or C)	UGU or UGC

Arginine (Arg or R)	CGU or CGC or CGA or CGG or AGA or AGG
Glycine (Gly or G)	GGU or GGC or GGA or GGG
Tryptophan (Trp or W)	UGG
Termination codon	UAA (ochre) or UAG (amber) or UGA (opal)

It should be understood that the codons specified above are for RNA sequences. The corresponding codons for DNA have a T substituted for U.

Mutations can be made in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39, such that a particular codon is changed to a codon which codes for a different amino acid. Such a mutation is generally made by making the fewest nucleotide changes possible. A substitution mutation of this sort can be made to change an amino acid in the resulting protein in a non-conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping of amino acids having a particular size or characteristic to an amino acid belonging to another grouping) or in a conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping of amino acids having a particular size or characteristic to an amino acid belonging to the same grouping). Such a conservative change generally leads to less change in the structure and function of the resulting protein. A non-conservative change is more likely to alter the structure, activity or function of the resulting protein. The present invention should be considered to include sequences containing conservative changes which do not significantly alter the activity or binding characteristics of the resulting protein.

The following is one example of various groupings of amino acids:

Amino acids with nonpolar R groups

Alanine

Valine

Leucine

Isoleucine

Proline

Phenylalanine

Tryptophan

Methionine

Amino acids with uncharged polar R groups

Glycine

Serine

Threonine

Cysteine

Tyrosine

Asparagine

Glutamine

Amino acids with charged polar R groups (negatively charged at pH 6.0)

Aspartic acid

Glutamic acid

Basic amino acids (positively charged at pH 6.0)

Lysine

Arginine

Histidine (at pH 6.0)

Another grouping may be those amino acids with phenyl groups:

Phenylalanine

Tryptophan

Tyrosine

Another grouping may be according to molecular weight (i.e., size of R groups):

Glycine	75
Alanine	89
Serine	105
Proline	115
Valine	117
Threonine	119
Cysteine	121
Leucine	131
Isoleucine	131
Asparagine	132
Aspartic acid	133
Glutamine	146
Lysine	146
Glutamic acid	147
Methionine	149
Histidine (at pH 6.0)	155
Phenylalanine	165
Arginine	174
Tyrosine	181
Tryptophan	204

Particularly preferred substitutions are:

- Lys for Arg and vice versa such that a positive charge may be maintained;
- Glu for Asp and vice versa such that a negative charge may be maintained;

- Ser for Thr such that a free -OH can be maintained; and
- Gln for Asn such that a free NH_2 can be maintained.

Amino acid substitutions may also be introduced to substitute an amino acid with a particularly preferable property. For example, a Cys may be introduced a potential site for disulfide bridges with another Cys. A His may be introduced as a particularly "catalytic" site (i.e., His can act as an acid or base and is the most common amino acid in biochemical catalysis). Pro may be introduced because of its particularly planar structure, which induces β-turns in the protein's structure.

Two amino acid sequences are "substantially homologous" when at least about 70% of the amino acid residues (preferably at least about 80%, and most preferably at least about 90 or 95%) are identical, or represent conservative substitutions.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., a cDNA where the genomic coding sequence contains introns, or synthetic sequences having codons different than the native gene). Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

An "antibody" is any immunoglobulin, including antibodies and fragments thereof, that binds a specific epitope. The term encompasses polyclonal, monoclonal, and chimeric antibodies, the last mentioned described in further detail in U.S. Patent Nos. 4,816,397 and 4,816,567.

An "antibody combining site" is that structural portion of an antibody molecule comprised of heavy and light chain variable and hypervariable regions that specifically binds antigen.

The phrase "antibody molecule" in its various grammatical forms as used herein contemplates both an intact immunoglobulin molecule and an immunologically active portion of an immunoglobulin molecule.

Exemplary antibody molecules are intact immunoglobulin molecules, substantially intact immunoglobulin molecules and those portions of an immunoglobulin molecule that contains the paratope, including those portions known in the art as Fab, Fab', $F(ab')_2$ and F(v), which portions are preferred for use in the therapeutic methods described herein.

Fab and $F(ab')_2$ portions of antibody molecules are prepared by the proteolytic reaction of papain and pepsin, respectively, on substantially intact antibody molecules by methods that are well-known. See for example, U.S. Patent No. 4,342,566 to Theofilopolous et al. Fab' antibody molecule portions are also well-known and are produced from $F(ab')_2$ portions followed by reduction of the disulfide bonds linking the two heavy chain portions as with mercaptoethanol, and followed by alkylation of the resulting protein mercaptan with a reagent such as iodoacetamide. An antibody containing intact antibody molecules is preferred herein.

The phrase "monoclonal antibody" in its various grammatical forms refers to an antibody having only one species of antibody combining site capable of immunoreacting with a particular antigen. A monoclonal antibody thus typically displays a single binding affinity for any antigen with which it immunoreacts. A monoclonal antibody may therefore contain an antibody molecule having a plurality

of antibody combining sites, each immunospecific for a different antigen; e.g., a bispecific (chimeric) monoclonal antibody.

The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that are physiologically tolerable and do not typically produce an allergic or similar untoward reaction, such as gastric upset, dizziness and the like, when administered to a human.

The phrase "therapeutically effective amount" is used herein to mean an amount sufficient to prevent, and preferably reduce by at least about 30 percent, more preferably by at least 50 percent, most preferably by at least 90 percent, a clinically significant change in the S phase activity of a target cellular mass, or other feature of pathology such as for example, elevated blood pressure, fever or white cell count as may attend its presence and activity.

A DNA sequence is "operatively linked" to an expression control sequence when the expression control sequence controls and regulates the transcription and translation of that DNA sequence. The term "operatively linked" includes having an appropriate start signal (e.g., ATG) in front of the DNA sequence to be expressed and maintaining the correct reading frame to permit expression of the DNA sequence under the control of the expression control sequence and production of the desired product encoded by the DNA sequence. If a gene that one desires to insert into a recombinant DNA molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene.

The term "standard hybridization conditions" refers to salt and temperature conditions substantially equivalent to 5 x SSC and 65°C for both hybridization and wash. However, one skilled in the art will appreciate that such "standard hybridization conditions" are dependent on particular conditions including the concentration of sodium and magnesium in the buffer, nucleotide sequence length

and concentration, percent mismatch, percent formamide, and the like. Also important in the determination of "standard hybridization conditions" is whether the two sequences hybridizing are RNA-RNA, DNA-DNA or RNA-DNA. Such standard hybridization conditions are easily determined by one skilled in the art according to well known formulae, wherein hybridization is typically 10-20°C below the predicted or determined T_m with washes of higher stringency, if desired.

In its primary aspect, the present invention concerns the identification of *cig* and *crg* genes and gene products and their use for the development of diagnostics, drug screening assays, and therapeutics for HCMV and other viral infections.

In a particular embodiment, the present invention relates to all members of the herein disclosed *cigs* and *crgs*.

The differential expression of the genes of this invention are diagnostic and characteristic of HCMV infection and interferon treatment. It is envisioned that these genes can be used as markers in assays designed to screen for compounds that are antagonistic to HCMV infection. The assays would utilize sequences that are complementary to the genes that are uniquely either induced or repressed upon HCMV infection as capture probes, attached individually to separate wells in a microtiter plate, or as an array on a flat solid support such as a nylon membrane, nitrocellulose membrane, glass sheet, or plastic sheet, in a hybridization-based assay. Measurement of the levels of expression from the different genes in infected cells, with or without treatment using test compounds, will reflect the efficacy of said compounds at either attenuating the expression of the HCMV-*i*nducible *g*enes (*cig*), or enhancing the expression of the HCMV-*r*epressed *g*enes (*crg*).

Measurement of expression levels will be facilitated by incorporating a detectable label into all newly synthesized RNAs post-HCMV infection or post-interferon treatment. These detectable labels, for example, radioactive- or fluorescent-labeled

ribonucleoside triphosphates, can be added immediately after infection or treatment, and thus be incorporated into any newly synthesized RNA molecule. Alternatively, the capture probe can be labeled with a compound that can be selectively detected upon hybridization to a target. For example a fluorescent label can be detected by fluorescence polarization. In another example, a label (radioactive, fluorescent, chemiluminescent, colorimetric, or enzymatic) can be detected by selective release into solution or retention on the solid support. The former can be accomplished using a nuclease that selectively cleaves the duplex (or heteroduplex in the case of a DNA capture probe and an RNA target), thus releasing the label into the solution phase for subsequent detection. The latter can be accomplished by use of a nuclease that will selectively cleave the single-stranded capture probe but leave the hybridized (duplex or heteroduplex) capture probe, and its attached label, protected and thus retained on the solid support for subsequent detection. In yet another example, antibodies which are specific for heteroduplexes (*i.e.* DNA capture probe hybridized to RNA target) can be used in a standard ELISA-type assay for detection.

The results from the assays, when used in a drug screening mode, will not only identify compounds that alter HCMV-characteristic expression patterns, but will also reveal what the specific targets are of the various effective compounds identified. The narrowed down list of candidate compounds derived from this first screening will then need to go through a second screening in a model system (either *in vitro* or *in vivo*) of HCMV infection to determine true efficacy.

A similar assay system can be used to follow the performance of HCMV-specific drugs in patients. This can be a valuable tool in monitoring the effectiveness of a patient's treatment regimen that ultimately can lead to tailoring the treatment to best fit the patient. Clearly, the system can be simplified be using a single probe that is diagnostic of the efficacy of the particular compound being used for treatment.

As stated above, the present invention also relates to a recombinant DNA molecule or cloned gene, or a degenerate variant thereof, which encodes a *cig* or *crg* gene product, or a fragment thereof, that possesses an amino acid sequence set forth in the Sequence Listing (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 27, 29, 31, 33, 35, and 37); preferably a nucleic acid molecule, in particular a recombinant DNA molecule or cloned gene, encoding the *cig* or *crg* gene product has a nucleotide sequence or is complementary to a DNA sequence shown in the Sequence Listing (SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39).

The possibilities both diagnostic and therapeutic that are raised by the existence of the *cigs* and *crgs*, derive from the fact that they are either selectively expressed or repressed in response to both HCMV infection and interferon treatment. As suggested earlier and elaborated further on herein, the present invention contemplates pharmaceutical intervention in the cascade of reactions in which the *cig* and *crg* gene products are implicated, to modulate the activity initiated by HCMV or other viral infection.

As discussed earlier, the *cig* and *crg* gene products or their binding partners or other ligands or agents exhibiting either mimicry or antagonism to the *cig* and *crg* gene products or control over their production, may be prepared in pharmaceutical compositions, with a suitable carrier and at a strength effective for administration by various means to a patient experiencing an adverse medical condition associated with HCMV or other viral infection for the treatment thereof. A variety of administrative techniques may be utilized, among them parenteral techniques such as subcutaneous, intravenous and intraperitoneal injections, catheterizations and the like. Average quantities of the *cig* or *crg* gene product or their subunits may vary and in particular should be based upon the recommendations and prescription of a qualified physician or veterinarian.

Also, antibodies including both polyclonal and monoclonal antibodies, and drugs that modulate the production or activity of the *cig* or *crg* gene products and/or their subunits may possess certain diagnostic applications and may for example, be utilized for the purpose of detecting and/or measuring conditions such as viral infection or the like. For example, the *cig* and *crg* gene products or their subunits may be used to produce both polyclonal and monoclonal antibodies to themselves in a variety of cellular media, by known techniques such as the hybridoma technique utilizing, for example, fused mouse spleen lymphocytes and myeloma cells. Likewise, small molecules that mimic or antagonize the activity(ies) of the *cig* or *crg* gene products of the invention may be discovered or synthesized, and may be used in diagnostic and/or therapeutic protocols.

The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal, antibody-producing cell lines can also be created by techniques other than fusion, such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. See, e.g., M. Schreier et al., "Hybridoma Techniques" (1980); Hammerling et al., "Monoclonal Antibodies And T-cell Hybridomas" (1981); Kennett et al., "Monoclonal Antibodies" (1980); see also U.S. Patent Nos. 4,341,761; 4,399,121; 4,427,783; 4,444,887; 4,451,570; 4,466,917; 4,472,500; 4,491,632; 4,493,890.

Panels of monoclonal antibodies produced against *cig* or *crg* gene product peptides can be screened for various properties; i.e., isotype, epitope, affinity, etc. Of particular interest are monoclonal antibodies that neutralize the activity of the *cig* gene products or their subunits. Such monoclonals can be readily identified in *cig* gene product activity assays. High affinity antibodies are also useful when immunoaffinity purification of native or recombinant *cig* or *crg* gene product is possible.

Preferably, the anti-*cig* or *crg* gene product antibody used in the diagnostic methods of this invention is an affinity purified polyclonal antibody. More preferably, the antibody is a monoclonal antibody (mAb). In addition, it is preferable for the anti-*cig* or *crg* gene product antibody molecules used herein be in the form of Fab, Fab', $F(ab')_2$ or F(v) portions of whole antibody molecules.

As suggested earlier, the diagnostic method of the present invention comprises examining a cellular sample or medium by means of an assay including an effective amount of an antagonist to a *cig* or *crg* gene product/protein, such as an anti-*cig* or *crg* gene product antibody, preferably an affinity-purified polyclonal antibody, and more preferably a mAb. In addition, it is preferable for the anti-*cig* or *crg* gene product antibody molecules used herein be in the form of Fab, Fab', $F(ab')_2$ or F(v) portions or whole antibody molecules. As previously discussed, patients capable of benefiting from this method include those suffering from viral infection (particularly with HCMV) or other like pathological derangement. Methods for isolating the *cig* or *crg* gene products and inducing anti-*cig* or *crg* gene product antibodies and for determining and optimizing the ability of anti-*cig* or *crg* gene product antibodies to assist in the examination of the target cells are all well-known in the art.

Methods for producing polyclonal anti-polypeptide antibodies are well-known in the art. See U.S. Patent No. 4,493,795 to Nestor et al. A monoclonal antibody, typically containing Fab and/or $F(ab')_2$ portions of useful antibody molecules, can be prepared using the hybridoma technology described in *Antibodies - A Laboratory Manual*, Harlow and Lane, eds., Cold Spring Harbor Laboratory, New York (1988), which is incorporated herein by reference. Briefly, to form the hybridoma from which the monoclonal antibody composition is produced, a myeloma or other self-perpetuating cell line is fused with lymphocytes obtained from the spleen of a mammal hyperimmunized with a *cig* or *crg* gene product-binding portion thereof, or *cig* or *crg* gene product, or an origin-specific DNA-binding portion thereof.

Splenocytes are typically fused with myeloma cells using polyethylene glycol (PEG) 6000. Fused hybrids are selected by their sensitivity to HAT. Hybridomas producing a monoclonal antibody useful in practicing this invention are identified by their ability to immunoreact with the present *cig* or *crg* gene product and their ability to inhibit specified *cig* or *crg* gene product activity in target cells.

A monoclonal antibody useful in practicing the present invention can be produced by initiating a monoclonal hybridoma culture comprising a nutrient medium containing a hybridoma that secretes antibody molecules of the appropriate antigen specificity. The culture is maintained under conditions and for a time period sufficient for the hybridoma to secrete the antibody molecules into the medium. The antibody-containing medium is then collected. The antibody molecules can then be further isolated by well-known techniques.

Media useful for the preparation of these compositions are both well-known in the art and commercially available and include synthetic culture media, inbred mice and the like. An exemplary synthetic medium is Dulbecco's minimal essential medium (DMEM; Dulbecco et al., *Virol.* **8**:396 (1959)) supplemented with 4.5 gm/l glucose, 20 mm glutamine, and 20% fetal calf serum. An exemplary inbred mouse strain is the Balb/c.

Methods for producing monoclonal anti-*cig* or *crg* gene product antibodies are also well-known in the art. See Niman et al., *Proc. Natl. Acad. Sci. USA*, **80**:4949-4953 (1983). Typically, the present *cig* or *crg* gene product or a peptide analog is used either alone or conjugated to an immunogenic carrier, as the immunogen in the before described procedure for producing anti-*cig* or *crg* gene product monoclonal antibodies. The hybridomas are screened for the ability to produce an antibody that immunoreacts with the *cig* or *crg* gene product peptide analog and the present *cig* or *crg* gene product.

The present invention further contemplates therapeutic compositions useful in practicing the therapeutic methods of this invention. A subject therapeutic composition includes, in admixture, a pharmaceutically acceptable excipient (carrier) and one or more of a *cig* or *crg* gene product, polypeptide analog thereof or fragment thereof, as described herein as an active ingredient. In a preferred embodiment, the composition comprises an antigen capable of modulating the specific binding of the present *cig* or *crg* gene product within a target cell.

The preparation of therapeutic compositions which contain polypeptides, analogs or active fragments as active ingredients is well understood in the art. Typically, such compositions are prepared as injectables, either as liquid solutions or suspensions, however, solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified. The active therapeutic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents which enhance the effectiveness of the active ingredient.

A polypeptide, analog or active fragment can be formulated into the therapeutic composition as neutralized pharmaceutically acceptable salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide or antibody molecule) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The therapeutic polypeptide-, analog- or active fragment-containing compositions are conventionally administered intravenously, as by injection of a unit dose, for example. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as unitary dosage for humans, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; i.e., carrier, or vehicle.

The compositions are administered in a manner compatible with the dosage formulation, and in a therapeutically effective amount. The quantity to be administered depends on the subject to be treated, capacity of the subject's immune system to utilize the active ingredient, and degree of inhibition or neutralization of binding capacity desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual. However, suitable dosages may range from about 0.1 to 20, preferably about 0.5 to about 10, and more preferably one to several, milligrams of active ingredient per kilogram body weight of individual per day and depend on the route of administration. Suitable regimes for initial administration and booster shots are also variable, but are typified by an initial administration followed by repeated doses at one or more hour intervals by a subsequent injection or other administration. Alternatively, continuous intravenous infusion sufficient to maintain concentrations of ten nanomolar to ten micromolar in the blood are contemplated.

The therapeutic compositions may further include an effective amount of the *cig* or *crg* gene product antagonist or analog thereof, and one or more of the following active ingredients: an antibiotic, a steroid. Exemplary formulations are given below:

<u>Formulations</u>

Intravenous Formulation I

Ingredient	mg/ml
cefotaxime	250.0
cig or *crg* gene product	10.0
dextrose USP	45.0
sodium bisulfite USP	3.2
edetate disodium USP	0.1
water for injection q.s.a.d.	1.0 ml

Intravenous Formulation II

Ingredient	mg/ml
ampicillin	250.0
cig or *crg* gene product	10.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml

Intravenous Formulation III

Ingredient	mg/ml
gentamicin (charged as sulfate)	40.0
cig or *crg* gene product	10.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml

Intravenous Formulation IV

Ingredient	mg/ml
cig or *crg* gene product	10.0
dextrose USP	45.0
sodium bisulfite USP	3.2

edetate disodium USP	0.1
water for injection q.s.a.d.	1.0 ml

<u>Intravenous Formulation V</u>

<u>Ingredient</u>	<u>mg/ml</u>
cig or *crg* gene product antagonist	5.0
sodium bisulfite USP	3.2
disodium edetate USP	0.1
water for injection q.s.a.d.	1.0 ml

As used herein, "pg" means picogram, "ng" means nanogram, "ug" or "μg" mean microgram, "mg" means milligram, "ul" or "μl" mean microliter, "ml" means milliliter, "l" means liter.

Another feature of this invention is the expression of the DNA sequences disclosed herein. As is well known in the art, DNA sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host.

Such operative linking of a DNA sequence of this invention to an expression control sequence, of course, includes, if not already part of the DNA sequence, the provision of an initiation codon, ATG, in the correct reading frame upstream of the DNA sequence.

A wide variety of host/expression vector combinations may be employed in expressing the DNA sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic DNA sequences. Suitable vectors include derivatives of SV40 and known bacterial plasmids, e.g., *E. coli* plasmids col El, pCR1, pBR322, pMB9 and their

derivatives, plasmids such as RP4; phage DNAS, e.g., the numerous derivatives of phage λ, e.g., NM989, and other phage DNA, e.g., M13 and filamentous single stranded phage DNA; yeast plasmids such as the 2μ plasmid or derivatives thereof; vectors useful in eukaryotic cells, such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or other expression control sequences; and the like.

Any of a wide variety of expression control sequences -- sequences that control the expression of a DNA sequence operatively linked to it -- may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include, for example, the early or late promoters of SV40, CMV, vaccinia, polyoma or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the *LTR* system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase (e.g., Pho5), the promoters of the yeast α-mating factors, and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

A wide variety of unicellular host cells are also useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, fungi such as yeasts, and animal cells, such as CHO, R1.1, B-W and L-M cells, African Green Monkey kidney cells (e.g., COS 1, COS 7, BSC1, BSC40, and BMT10), insect cells (e.g., Sf9), and human cells and plant cells in tissue culture.

It will be understood that not all vectors, expression control sequences and hosts will function equally well to express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one

skilled in the art will be able to select the proper vectors, expression control sequences, and hosts without undue experimentation to accomplish the desired expression without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must function in it. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, will also be considered.

In selecting an expression control sequence, a variety of factors will normally be considered. These include, for example, the relative strength of the system, its controllability, and its compatibility with the particular DNA sequence or gene to be expressed, particularly as regards potential secondary structures. Suitable unicellular hosts will be selected by consideration of, e.g., their compatibility with the chosen vector, their secretion characteristics, their ability to fold proteins correctly, and their fermentation requirements, as well as the toxicity to the host of the product encoded by the DNA sequences to be expressed, and the ease of purification of the expression products.

Considering these and other factors a person skilled in the art will be able to construct a variety of vector/expression control sequence/host combinations that will express the DNA sequences of this invention on fermentation or in large scale animal culture.

It is further intended that *cig* or *crg* gene product analogs may be prepared from nucleotide sequences of the protein complex/subunit derived within the scope of the present invention. Analogs, such as fragments, may be produced, for example, by pepsin digestion of *cig* or *crg* gene product material. Other analogs, such as muteins, can be produced by standard site-directed mutagenesis of *cig* or *crg* gene product coding sequences. Analogs exhibiting "*cig* or *crg* gene product activity"

such as small molecules, whether functioning as promoters or inhibitors, may be identified by known *in vivo* and/or *in vitro* assays.

As mentioned above, a DNA sequence encoding *cig* or *crg* gene product can be prepared synthetically rather than cloned. The DNA sequence can be designed with the appropriate codons for the *cig* or *crg* gene product amino acid sequence. In general, one will select preferred codons for the intended host if the sequence will be used for expression. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature*, **292**:756 (1981); Nambair et al., *Science*, **223**:1299 (1984); Jay et al., *J. Biol. Chem.*, *259*:6311 (1984).

Synthetic DNA sequences allow convenient construction of genes which will express *cig* or *crg* gene product analogs or "muteins". Alternatively, DNA encoding muteins can be made by site-directed mutagenesis of native *cig* or *crg* gene product genes or cDNAs, and muteins can be made directly using conventional polypeptide synthesis.

A general method for site-specific incorporation of unnatural amino acids into proteins is described in Christopher J. Noren, Spencer J. Anthony-Cahill, Michael C. Griffith, Peter G. Schultz, *Science*, **244**:182-188 (April 1989). This method may be used to create analogs with unnatural amino acids.

The present invention extends to the preparation of antisense oligonucleotides and ribozymes that may be used to interfere with the expression of the ˜ at the translational level. This approach utilizes antisense nucleic acid and ribozymes to block translation of a specific mRNA, either by masking that mRNA with an antisense nucleic acid or cleaving it with a ribozyme.

Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule. (See Weintraub, 1990; Marcus-Sekura, 1988.) In the cell, they hybridize to that mRNA, forming a double stranded molecule. The cell does not translate an mRNA in this double-stranded form. Therefore, antisense nucleic acids interfere with the expression of mRNA into protein. Oligomers of about fifteen nucleotides and molecules that hybridize to the AUG initiation codon will be particularly efficient, since they are easy to synthesize and are likely to pose fewer problems than larger molecules when introducing them into ˜-producing cells. Antisense methods have been used to inhibit the expression of many genes *in vitro* (Marcus-Sekura, 1988; Hambor et al., 1988).

Ribozymes are RNA molecules possessing the ability to specifically cleave other single stranded RNA molecules in a manner somewhat analogous to DNA restriction endonucleases. Ribozymes were discovered from the observation that certain mRNAs have the ability to excise their own introns. By modifying the nucleotide sequence of these RNAs, researchers have been able to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, 1988.). Because they are sequence-specific, only mRNAs with particular sequences are inactivated.

Investigators have identified two types of ribozymes, *Tetrahymena*-type and "hammerhead"-type. (Hasselhoff and Gerlach, 1988) *Tetrahymena*-type ribozymes recognize four-base sequences, while "hammerhead"-type recognize eleven- to eighteen-base sequences. The longer the recognition sequence, the more likely it is to occur exclusively in the target mRNA species. Therefore, hammerhead-type ribozymes are preferable to *Tetrahymena*-type ribozymes for inactivating a specific mRNA species, and eighteen base recognition sequences are preferable to shorter recognition sequences.

The DNA sequences described herein may thus be used to prepare antisense molecules against, and ribozymes that cleave mRNAs for *cig* or *crg* gene product and their ligands.

In one embodiment, a gene encoding a *cig* or *crg* gene product or polypeptide domain fragment thereof is introduced *in vivo* in a viral vector. Such vectors include an attenuated or defective DNA virus, such as but not limited to herpes simplex virus (HSV), papillomavirus, Epstein Barr virus (EBV), adenovirus, adeno-associated virus (AAV), and the like. Defective viruses, which entirely or almost entirely lack viral genes, are preferred. Defective virus is not infective after introduction into a cell. Use of defective viral vectors allows for administration to cells in a specific, localized area, without concern that the vector can infect other cells. Examples of particular vectors include, but are not limited to, a defective herpes virus-1 (HSV-1) vector [Kaplitt et al., *Molec. Cell. Neurosci.* **2**:320-330 (1991)], an attenuated adenovirus vector, such as the vector described by Stratford-Perricaudet et al. [*J. Clin. Invest.* **90**:626-630 (1992)], and a defective adeno-associated virus vector [Samulski et al., *J. Virol.* **61**:3096-3101 (1987); Samulski et al., *J. Virol.* **63**:3822-3828 (1989)].

Preferably, for *in vitro* administration, an appropriate immunosuppressive treatment is employed in conjunction with the viral vector, *e.g.*, adenovirus vector, to avoid immuno-deactivation of the viral vector and transfected cells. For example, immunosuppressive cytokines, such as interleukin-12 (IL-12), interferon-γ (IFN-γ), or anti-CD4 antibody, can be administered to block humoral or cellular immune responses to the viral vectors [*see, e.g.*, Wilson, *Nature Medicine* (1995)]. In addition, it is advantageous to employ a viral vector that is engineered to express a minimal number of antigens.

In another embodiment the gene can be introduced in a retroviral vector, *e.g.*, as described in Anderson et al., U.S. Patent No. 5,399,346; Mann et al., 1983, Cell

33:153; Temin et al., U.S. Patent No. 4,650,764; Temin et al., U.S. Patent No. 4,980,289; Markowitz et al., 1988, J. Virol. 62:1120; Temin et al., U.S. Patent No. 5,124,263; International Patent Publication No. WO 95/07358, published March 16, 1995, by Dougherty et al.; and Kuo et al., 1993, Blood 82:845.

Targeted gene delivery is described in International Patent Publication WO 95/28494, published October 1995.

Alternatively, the vector can be introduced *in vivo* by lipofection. For the past decade, there has been increasing use of liposomes for encapsulation and transfection of nucleic acids *in vitro*. Synthetic cationic lipids designed to limit the difficulties and dangers encountered with liposome mediated transfection can be used to prepare liposomes for *in vivo* transfection of a gene encoding a marker [Felgner, et. al., *Proc. Natl. Acad. Sci. U.S.A.* **84**:7413-7417 (1987); *see* Mackey, et al., *Proc. Natl. Acad. Sci. U.S.A.* **85**:8027-8031 (1988)]. The use of cationic lipids may promote encapsulation of negatively charged nucleic acids, and also promote fusion with negatively charged cell membranes [Felgner and Ringold, *Science* **337**:387-388 (1989)]. The use of lipofection to introduce exogenous genes into the specific organs *in vivo* has certain practical advantages. Molecular targeting of liposomes to specific cells represents one area of benefit. It is clear that directing transfection to particular cell types would be particularly advantageous in a tissue with cellular heterogeneity, such as pancreas, liver, kidney, and the brain. Lipids may be chemically coupled to other molecules for the purpose of targeting [*see* Mackey, et. al., *supra*]. Targeted peptides, *e.g.*, hormones or neurotransmitters, and proteins such as antibodies, or non-peptide molecules could be coupled to liposomes chemically.

It is also possible to introduce the vector *in vivo* as a naked DNA plasmid. Naked DNA vectors for gene therapy can be introduced into the desired host cells by methods known in the art, *e.g.*, transfection, electroporation, microinjection,

transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun, or use of a DNA vector transporter [*see, e.g.*, Wu et al., *J. Biol. Chem.* **267**:963-967 (1992); Wu and Wu, *J. Biol. Chem.* **263**:14621-14624 (1988); Hartmut et al., Canadian Patent Application No. 2,012,311, filed March 15, 1990].

In a preferred embodiment of the present invention, a gene therapy vector as described above employs a transcription control sequence operably associated with the *cig* or *crg* sequence inserted in the vector. That is, a specific expression vector of the present invention can be used in gene therapy.

Such an expression vector is particularly useful to regulate expression of a therapeutic *cig* or *crg*. In one embodiment, the present invention contemplates constitutive expression of the *cig* or *crg*, even if at low levels. Various therapeutic heterologous genes can be inserted in a gene therapy vector of the invention such as but not limited to adenosine deaminase (ADA) to treat severe combined immunodeficiency (SCID); marker genes or lymphokine genes into tumor infiltrating (TIL) T cells [Kasis et al., *Proc. Natl. Acad. Sci. U.S.A.* **87**:473 (1990); Culver et al., *ibid.* **88**:3155 (1991)]; genes for clotting factors such as Factor VIII and Factor IX for treating hemophilia [Dwarki et al. *Proc. Natl. Acad. Sci. USA*, **92**:1023-1027 (19950); Thompson, *Thromb. and Haemostatis*, **66**:119-122 (1991)]; and various other well known therapeutic genes such as, but not limited to, β-globin, dystrophin, insulin, erythropoietin, growth hormone, glucocerebrosidase, β-glucuronidase, α-antitrypsin, phenylalanine hydroxylase, tyrosine hydroxylase, ornithine transcarbamylase, apolipoproteins, and the like. In general, see U.S. Patent No. 5,399,346 to Anderson et al.

In another aspect, the present invention provides for regulated expression of the heterologous gene in concert with expression of proteins under control of *** upon commitment to DNA synthesis. Concerted control of such heterologous genes may be particularly useful in the context of treatment for proliferative disorders, such as

tumors and cancers, when the heterologous gene encodes a targeting marker or immunomodulatory cytokine that enhances targeting of the tumor cell by host immune system mechanisms. Examples of such heterologous genes for immunomodulatory (or immuno-effector) molecules include, but are not limited to, interferon-α, interferon-γ, interferon-β, interferon-ω, interferon-τ, tumor necrosis factor-α, tumor necrosis factor-β, interleukin-2, interleukin-7, interleukin-12, interleukin-15, B7-1 T cell co-stimulatory molecule, B7-2 T cell co-stimulatory molecule, immune cell adhesion molecule (ICAM) -1 T cell co-stimulatory molecule, granulocyte colony stimulatory factor, granulocyte-macrophage colony stimulatory factor, and combinations thereof.

In a further embodiment, the present invention provides for co-expression of *cig* or *crg* and a therapeutic heterologous gene under control of a specific DNA recognition sequence by providing a gene therapy expression vector comprising both a *cig* or *crg* coding gene and a gene under control of, *inter alia*, the *cig* or *crg* regulatory sequence. In one embodiment, these elements are provided on separate vectors, *e.g.*, as exemplified *infra*. These elements may be provided in a single expression vector.

The present invention also relates to a variety of diagnostic applications, including methods for detecting the presence of stimuli such as the earlier referenced polypeptide ligands, by reference to their ability to elicit the activities which are mediated by the present *cig* or *crg* gene products. As mentioned earlier, the *cig* or *crg* gene products can be used to produce antibodies to itself by a variety of known techniques, and such antibodies could then be isolated and utilized as in tests for the presence of particular *cig* or *crg* gene product activity in suspect target cells.

As described in detail above, antibody(ies) to the *cig* or *crg* gene products can be produced and isolated by standard methods including the well known hybridoma

techniques. For convenience, the antibody(ies) to the *cig* or *crg* gene products will be referred to herein as Ab_1 and antibody(ies) raised in another species as Ab_2.

The presence of ~ in cells can be ascertained by the usual immunological procedures applicable to such determinations. A number of useful procedures are known. Three such procedures which are especially useful utilize either the *cig* or *crg* gene product labeled with a detectable label, antibody Ab_1 labeled with a detectable label, or antibody Ab_2 labeled with a detectable label. The procedures may be summarized by the following equations wherein the asterisk indicates that the particle is labeled, and "~" stands for the *cig* or *crg* gene product:

A. $\sim^{*} + Ab_1 = \sim^{*}Ab_1$

B. $\sim + Ab^{*} = \sim Ab_1^{*}$

C. $\sim + Ab_1 + Ab_2^{*} = \sim Ab_1Ab_2^{*}$

The procedures and their application are all familiar to those skilled in the art and accordingly may be utilized within the scope of the present invention. The "competitive" procedure, Procedure A, is described in U.S. Patent Nos. 3,654,090 and 3,850,752. Procedure C, the "sandwich" procedure, is described in U.S. Patent Nos. RE 31,006 and 4,016,043. Still other procedures are known such as the "double antibody," or "DASP" procedure.

In each instance, the *cig* or *crg* gene product forms complexes with one or more antibody(ies) or binding partners and one member of the complex is labeled with a detectable label. The fact that a complex has formed and, if desired, the amount thereof, can be determined by known methods applicable to the detection of labels.

It will be seen from the above, that a characteristic property of Ab_2 is that it will react with Ab_1. This is because Ab_1 raised in one mammalian species has been used in another species as an antigen to raise the antibody Ab_2. For example, Ab_2 may be raised in goats using rabbit antibodies as antigens. Ab_2 therefore would be

anti-rabbit antibody raised in goats. For purposes of this description and claims, Ab_1 will be referred to as a primary or anti-*cig* or *crg* gene product antibody, and Ab_2 will be referred to as a secondary or anti-Ab_1 antibody.

The labels most commonly employed for these studies are radioactive elements, enzymes, chemicals which fluoresce when exposed to ultraviolet light, and others.

A number of fluorescent materials are known and can be utilized as labels. These include, for example, fluorescein, rhodamine, auramine, Texas Red, AMCA blue and Lucifer Yellow. A particular detecting material is anti-rabbit antibody prepared in goats and conjugated with fluorescein through an isothiocyanate.

The *cig* or *crg* gene product or its binding partner(s) can also be labeled with a radioactive element or with an enzyme. The radioactive label can be detected by any of the currently available counting procedures. The preferred isotope may be selected from ^{3}H, ^{14}C, ^{32}P, ^{33}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re.

Enzyme labels are likewise useful, and can be detected by any of the presently utilized colorimetric, spectrophotometric, fluorospectrophotometric, amperometric or gasometric techniques. The enzyme is conjugated to the selected particle by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde and the like. Many enzymes which can be used in these procedures are known and can be utilized. The preferred are peroxidase, ß-glucuronidase, ß-D-glucosidase, ß-D-galactosidase, urease, glucose oxidase plus peroxidase and alkaline phosphatase. U.S. Patent Nos. 3,654,090; 3,850,752; and 4,016,043 are referred to by way of example for their disclosure of alternate labeling material and methods.

A particular assay system developed and utilized in accordance with the present invention, is known as a receptor assay. In a receptor assay, the material to be assayed is appropriately labeled and then certain cellular test colonies are inoculated with a quantity of both the labeled and unlabeled material after which binding studies are conducted to determine the extent to which the labeled material binds to the cell receptors. In this way, differences in affinity between materials can be ascertained.

Accordingly, a purified quantity of the *cig* or *crg* gene product may be radiolabeled and combined, for example, with antibodies or other inhibitors thereto, after which binding studies would be carried out. Solutions would then be prepared that contain various quantities of labeled and unlabeled uncombined *cig* or *crg* gene product, and cell samples would then be inoculated and thereafter incubated. The resulting cell monolayers are then washed, solubilized and then counted in a gamma counter for a length of time sufficient to yield a standard error of <5%. These data are then subjected to Scatchard analysis after which observations and conclusions regarding material activity can be drawn. While the foregoing is exemplary, it illustrates the manner in which a receptor assay may be performed and utilized, in the instance where the cellular binding ability of the assayed material may serve as a distinguishing characteristic.

An assay useful and contemplated in accordance with the present invention is known as a "cis/trans" assay. Briefly, this assay employs two genetic constructs, one of which is typically a plasmid that continually expresses a particular receptor of interest when transfected into an appropriate cell line, and the second of which is a plasmid that expresses a reporter such as luciferase, under the control of a receptor/ligand complex. Thus, for example, if it is desired to evaluate a compound as a ligand for a particular receptor, one of the plasmids would be a construct that results in expression of the receptor in the chosen cell line, while the second plasmid would possess a promoter linked to the luciferase gene in which the

response element to the particular receptor is inserted. If the compound under test is an agonist for the receptor, the ligand will complex with the receptor, and the resulting complex will bind the response element and initiate transcription of the luciferase gene. The resulting chemiluminescence is then measured photometrically, and dose response curves are obtained and compared to those of known ligands. The foregoing protocol is described in detail in U.S. Patent No. 4,981,784 and PCT International Publication No. WO 88/03168, for which purpose the artisan is referred.

In a further embodiment of this invention, commercial test kits suitable for use by a medical specialist may be prepared to determine the presence or absence of predetermined *cig* or *crg* gene product activity in suspected target cells. In accordance with the testing techniques discussed above, one class of such kits will contain at least the labeled *cig* or *crg* gene product or its binding partner, for instance an antibody specific thereto, and directions, of course, depending upon the method selected, e.g., "competitive," "sandwich," "DASP" and the like. The kits may also contain peripheral reagents such as buffers, stabilizers, etc.

Accordingly, a test kit may be prepared for the demonstration of the presence or capability of cells for predetermined *cig* or *crg* gene product activity, comprising:

(a) a predetermined amount of at least one labeled immunochemically reactive component obtained by the direct or indirect attachment of the present *cig* or *crg* gene product factor or a specific binding partner thereto, to a detectable label;

(b) other reagents; and

(c) directions for use of said kit.

More specifically, the diagnostic test kit may comprise:

(a) a known amount of the *cig* or *crg* gene products as described above (or a binding partner) generally bound to a solid phase to form an immunosorbent, or in

the alternative, bound to a suitable tag, or plural such end products, etc. (or their binding partners) one of each;

(b) if necessary, other reagents; and

(c) directions for use of said test kit.

In a further variation, the test kit may be prepared and used for the purposes stated above, which operates according to a predetermined protocol (e.g. "competitive," "sandwich," "double antibody," etc.), and comprises:

(a) a labeled component which has been obtained by coupling the ˜ to a detectable label;

(b) one or more additional immunochemical reagents of which at least one reagent is a ligand or an immobilized ligand, which ligand is selected from the group consisting of:

(i) a ligand capable of binding with the labeled component (a);

(ii) a ligand capable of binding with a binding partner of the labeled component (a);

(iii) a ligand capable of binding with at least one of the component(s) to be determined; and

(iv) a ligand capable of binding with at least one of the binding partners of at least one of the component(s) to be determined; and

(c) directions for the performance of a protocol for the detection and/or determination of one or more components of an immunochemical reaction between the ˜ and a specific binding partner thereto.

In accordance with the above, an assay system for screening potential drugs effective to modulate the activity of the *cig* or *crg* gene product may be prepared. The *cig* or *crg* gene product may be introduced into a test system, and the prospective drug may also be introduced into the resulting cell culture, and the culture thereafter examined to observe any changes in the *cig* or *crg* gene products

activity of the cells, due either to the addition of the prospective drug alone, or due to the effect of added quantities of the known *cig* or *crg* gene product.

The following examples are presented in order to more fully illustrate the preferred embodiments of the invention. They should in no way be construed, however, as limiting the broad scope of the invention.

EXAMPLE 1

Cells and viruses. Primary human foreskin (HF) cells were cultures in medium containing 10% fetal calf serum. Cells were held at confluence for 3-4 days prior to experimentation. To avoid cell stimulation by fresh serum, treated cells were returned to the medium in which they were previously maintained. Where indicated, HF cells were treated with 500U/ml interferon-α and β (Sigman) for 4 h, and 100 μg/ml cyclohexamide was used to block protein synthesis.

HF cells were infected with HCMV strain AD169 (18), Towne (19) or Toledo (20). Wild-type adenovirus, *dl*309 (21), and herpes simplex virus type 1 (HSV-1) were also used. Infections with HCMV or HSV-1 were performed at a multiplicity of 3 plaque-forming units/cell, and adenovirus was used at a multiplicity of 30 plaque-forming units/cell. For inactivation with UV light, 5 ml medium containing HCMV was placed in a 15-cm-diameter dish, and irradiated at 2J/m2/sec for 10 min with mixing every 2 min. UV-treated stocks failed to produce detectable IE1 and IE2 protein at 8 or 36 h after infection. For neutralization, 50 μl HCMV stock was incubated with 20 μl neutralizing antibody (gift from Jay Nelson, University of Oregon) for 1 h at room temperature. Neutralization was confirmed by plaque assay. HCMV particles were concentrated and purified as described previously (22). HCMV membrane and tegument/capsid proteins were separated and isolated by detergent stripping (23).

EXAMPLE 2

Differential display assay. For differential display analysis (16-17), HF cells were mock-infected or infected with AD169 or UV-inactivated AD169. Total RNA was isolated 8 h later by using the TRIZOL Reagent (Life Technologies). First-strand cDNAs were synthesized using oligo(dT), and amplified in parallel PCR reactions in the presence of [α-^{33}P]dCTP using 135 combinations of 19 primers (Delta RNA Fingerprinting Kit, Clontech). The products were separated by electrophoresis on 5% polyacrylamide gels containing 8M urea. Differentially expressed bands were cut out of the gel, reamplified using the appropriate primer set, cloned into the pT7Blue T-Vector (Novagen), sequenced, and the results were analyzed by BLAST search (National Center for Biotechnology Information).

EXAMPLE 3

Assays for RNAs and proteins. For Northern blot assays 5 μg RNA from mock- or HCMV-infected HF cells was probed with random hexanucleotide-primed ^{32}P-labeled cDNA clones. The probes for mxA, isg15K and interferon-β were the partial cDNA sequences purified from I.M.A.G.E. Consortium (LLNL) clones (Genome Systems). For Western blot assays, three mouse monoclonal antibodies that recognize HCMV proteins, anti-IE1/IE2 (MAb810, Chemicon), anti-pp65 (2) and anti-glycoprotein B (Goodwin Institute), were used as the primary antibodies. Mab810 and anti-pp65 were also used for immunofluorescent staining.

EXAMPLE 4

Analysis of Cytomegalovirus-Induced RNAs. HCMV could alter host cell gene expression through the action of virion proteins or by the synthesis of new viral proteins after infection. To distinguish between these possibilities, we compared competent virus (HCMV) to UV-inactivated virus (UV HCMV). To test the effect of UV treatment, the delivery of the pp65 virion protein to the cells and the synthesis of the IE1 and IE2 immediate-early proteins were monitored. UV irradiation did not affect viral entry into the cells because the amount of pp65 delivered to the cells did not change with UV treatment (Fig. 1A). The IE1 and

IE2 proteins were detected at 8 and 21 h after infection in HCMV-infected cells, but not in UV HCMV-infected cells (Fig. 1A). Inhibition of viral RNA accumulation in UV HCMV-infected cells was also evident. The IE1 transcript could be detected at 8 h after infection in HCMV-infected cells, but not in UV HCMV-infected cells (Fig. 1B). We also determined the location of a virion protein in cells infected with UV-treated virus. pp65 was visible in nuclei at 2 h after infection with either HCMV or UV HCMV (Fig. 1C, panel 1 and 2). As expected, IE1 protein was detected in HCMV-infected but not in UV HCMV-infected cells (Fig. 1C, panel 3 and 4). These experiments demonstrate that UV irradiation of virus particles blocked the accumulation of detectable amounts of HCMV-encoded RNA without preventing the entrance of the virus into the cell or altering the intracellular localization of a virion protein.

We compared RNA levels by differential display (16, 17) at 8 h after infection or mock infection. HCMV immediate-early proteins have accumulated to significant levels at this time (see Fig. 5), giving them an opportunity to influence host cell mRNA accumulation. PCR-generated bands that were evident in virus-infected but not mock-infected samples could be divided into two groups. One group contained an induced band that was present in the HCMV-infected sample, but not in the UV HCMV-infected sample. The induced bands in this group could be derived from either viral or cellular RNAs. The second group contained induced bands in both HCMV- and UV HCMV-infected samples. These bands should represent cellular RNAs that accumulate after HCMV infection, since viral mRNAs are not produced in UV HCMV-infected cells (Fig. 1B).

We selected 71 of the most strongly induced PCR-generated bands for analysis. These DNA fragments were reamplified by PCR, cloned, and used as probes for Northern blot analyses to confirm that the bands represented differentially expressed genes. Examples of these assays are displayed in Figure 2A. Most of the cloned cDNA segments identified RNAs that were present at very low or non-

detectable levels in mock-infected cells, but accumulated to a high level in infected cells. cDNA clones representing up-regulated RNAs were isolated from 57 of the 71 reamplified fragments. Each clone is termed a *cig* for *C*MV *i*nducible *g*ene.

Thirty of 57 *cig* RNAs were induced by HCMV but not UV HCMV infection, and sequence analysis revealed that all of these clones corresponded to viral RNAs (data not shown). Two of the viral RNAs were produced after infection in the presence of cycloheximide identifying them as immediate-early RNAs, and the synthesis of the remainder was inhibited by the drug, indicating that they are early RNAs (Fig. 2B and data not shown).

Infection with either HCMV or UV HCMV led to the accumulation of 27 of the 57 *cig* RNAs, and sequence analysis demonstrated that they correspond to as many as 15 different cellular genes (Table 1). Nine were previously identified, and the other 6 were not found in a BLAST search. Surprisingly, most of the known RNAs previously were shown to be induced by interferon-α in HF cells, as were the 6 new RNAs (Fig. 2C and data not shown). The RNAs were induced both by virus infection and interferon-α in three lots of HF cells derived from different individuals (data not shown). Since the RNAs induced by infection corresponded to interferon-inducible genes, it seemed possible that other interferon-stimulated genes might be induced by HCMV. As expected, RNAs corresponding to mxA (33, 34), ISG15K (35, 36) and interferon-β (37) also were induced (Fig. 2C). As controls, we tested the expression of p53, p21, cytosolic phospholipase A2 (cPLA2) and actin. The level of these RNAs did not change after infection (Fig. 2D and Fig. 5).

Table 1. Cellular cDNA clones identified by differential display analysis

Clone	Gene	Reference
cig 1, 22, 51	interferon-stimulated gene 54K	24
cig 19	KIAA0062	25
cig 24, 70	glyceraldehyde-3-phosphate dehydrogenase	26
cig 25	guanylate binding protein isoform I	27
cig 32	Mn-superoxide dismutase	28
cig 34, 45, 46, 68	microtubular aggregate protein, p44	29
cig 43	IFP53	30
cig 52	(2'-5') oligoadenylate synthetase	31
cig 53	guanylate binding protein isoform II	32
cig 5-7, 15, 18, 44, 61, 69	new	this patent
cig 33	new	this patent
cig 41	new	this patent
cig 42	new	this patent
cig 49	new	this patent
cig 64	new	this patent

EXAMPLE 5

HCMV particles induce the accumulation of *cig* RNAs encoded by cellular genes. The differential display analysis utilized the laboratory adapted AD169 strain of HCMV. Towne, a second laboratory adapted HCMV strain, and Toledo, a low passage clinical isolate of HCMV, also strongly activated the accumulation of cell-coded *cig* RNAs (Fig. 3A, lane 10 and 11). Wild-type adenovirus did not activate the accumulation of *cig* RNAs and HSV-1 increased their expression to a very limited extent (Fig. 3A, lanes 8 and 9; Fig. 5). The expression of an adenovirus and HSV-1 mRNA was monitored to be certain that cells were

successfully infected (data not shown). Thus, whereas multiple HCMV strains strongly induced *cig* RNA accumulation, two other viruses did not.

To ask if cellular protein synthesis was required for the induction of cellular interferon-responsive RNAs, cells were infected in the presence of cycloheximide. It did not block the induction of *cig* RNAs by HCMV, and the drug itself had no effect on *cig* RNA expression (Fig. 3A, lane 4 and 5). This result indicates that the accumulation of *cig* RNAs does not require the synthesis of viral or cellular proteins after infection. It also rules out the possibility that a protein factor, such as a cytokine, is synthesized in response to the infection, and released from the cell so that it can interact with a cell surface receptor to induce *cig* RNAs.

Because infected cell lysates were used as virus stocks in our initial experiments, it was possible that soluble signaling molecules were present that could mediate the induction of RNAs encoded by the cell. We therefore performed a series of experiments to identify the component in HCMV stocks that was responsible for the induction. Initially, an HCMV stock was separated into two fractions by filtration through a 100 kDa cutoff membrane. The virus fraction was further purified by rate-velocity centrifugation, separating infectious virions and non-infectious enveloped particles (NIEPs, lacking viral DNA). The filtered lysate, purified virions and NIEPs were used to treat cells, and their abilities to induce the accumulation of *cig* RNAs were assayed. Purified virions and NIEPs activated *cig* RNA accumulation (Fig. 3A, lane 6 and 7), while the filtered lysate had little effect (Fig. 3A, lane 3). To prove that small molecules could pass through the filter, interferon-α (500 U/ml) was added to the infected cell lysate, and there was no loss of interferon activity after filtration (Fig 3A, lane 13 and 14).

We used neutralizing antibodies to confirm our observation indicating that the activation of *cig* RNA accumulation is mediated by HCMV particles and not by interferon. When the virus stock was incubated with antibody to virions, its ability

to induce *cig* RNAs was blocked, while antibody to interferon-α or β had no effect (Fig. 3B). The same amounts of interferon-specific antibodies were sufficient to block interferon-α or β activity in uninfected cultures (Fig 3B). We conclude that the HCMV particle or a molecule tightly associated with the particle initiates the induction of cellular *cig* RNAs. Expression of viral gnes is not required, since purified NIEPs and UV HCMV can induce *cig* RNAs.

We next explored the possibility that interferon might be carried within the HCMV particle. Purified viral particles were treated with Triton X-100 (0.5%) and deoxycholate (0.5%) and subjected to centrifugation to produce a supernatant fraction containing HCMV membrane proteins and a pellet containing internal virion constituents. With detergent treatment, pp65 (a marker for the tegument/capsid fraction) was in the pellet fraction and gB (a marker for the membrane fraction) was in the supernatant fraction. Without detergent treatment, the particle remained intact, and both pp65 and gB were in the pellet fraction (data not shown). As expected, without detergent treatment, the pellet fraction, but not supernatant fraction, activated *cig* RNA accumulation; with detergent treatment, neither the pellet fraction, nor supernatant the fraction activated the accumulation (Fig. 4A). When interferon-α was treated with the detergent mixture, its activity was not affected (Fig. 4A). This experiment indicates that the intact virus particle is required for the induction of *cig* RNAs, and further argues that this induction is not due to contaminating interferons.

Our results argue that the induction of cell-coded *cig* RNAs does not result from contaminants in HCMV preparations or from newly synthesized signaling proteins. Nevertheless, one might propose that a trace amount of a signaling molecule is stored in the cell, secreted after infection, and then acts at the surface of neighboring cells to induce *cig* RNAs. Accordingly, we performed an experiment in which uninfected cells and cells infected 1 h earlier were mixed in a ratio of 9:1, and a sufficient number of cells were plated to generate a confluent monolayer. At

the same time, 100% infected cells or 100% non-infected cells were plated at the same density. RNA was prepared at 8 h after infection, and the expression of *cig* RNAs and the HCMV IE1 RNA were assayed. The viral and *cig* RNAs were induced in the infected culture, but not in the uninfected culture (Fig. 4B). The RNA levels were induced to the same extent in the mixed culture as was seen for an uninfected/infected (ratio, 9:1) cell mixture prepared immediately before the extraction of RNA (Fig. 4B). Infected cells did not significantly induce the accumulation of *cig* RNAs in their uninfected neighbors.

EXAMPLE 6

Kinetics of *cig* RNA induction by HCMV as compared to interferon-α. The kinetics of *cig* RNA accumulation varied when cells were treated with different inducers (Fig. 5). Accumulation was first evident at 4-6 h after infection with HCMV, *cig* RNA levels peaked at about 8 h, and remained at high levels for the duration of the experiment (48 h). The HCMV IE1 gene showed a similar expression pattern. The induction of *cig* RNA expression in cells treated with interferon-α was more rapid and transient. The *cig* RNAs were detected at 30 min and reached their peak at 2-4 h before declining rapidly. The marked difference in the kinetics of *cig* RNA accumulation in HCMV-infected as compared to interferon-treated cells further supports the conclusion that the induction observed subsequent to HCMV infection is not the result of contaminating interferon in virus preparations.

In HSV-1-infected cells, the induction of *cig* RNAs was very limited (Fig. 5), consistent with the view that the strong induction of *cig* RNA accumulation observed in HCMV-infected cells is not a common cellular response to all herpesviruses. As a control, the HSV-1 icp47 immediate-early gene was shown to be expressed at a high level, demonstrating that the culture was successfully infected.

Discussion

We cloned 57 partial cDNA segments corresponding to RNAs that are present at a higher concentration in HCMV-infected as compared to mock-infected human fibroblasts. The 57 clones represent no more than 26 different mRNAs because some of the RNAs corresponded to more than one cDNA fragment generated by different primer sets. It is possible that we have identified fewer than 26 distinct RNAs since 6 of the partial cellular cDNAs were not found in a BLAST search, and we have determined the complete sequence of only one of the newly discovered RNAs. Since the others are only partially sequenced, more than 1 of the remaining 5 sequences might be contained within the same RNA molecule. However, only 2 of the 5 partially sequenced clones appear to recognize RNAs of identical size in Northern blot assays (Fig. 2 and data not shown).

Of the 26 cDNA clones, 11 were virus-coded. All of the immediate-early and some early HCMV mRNAs should have accumulated to detectable levels at 8 h after infection when cells were harvested; and partial cDNA clones corresponding to both classes of viral RNA were isolated. The screen identified 2 from a total of approximately 10 immediate-early mRNAs. One can not accurately estimate the total number of HCMV early mRNAs expressed at 8 hr since the number increases continually from about 4 h to 24 h after infection (15). Given the uncertainties about the number of different viral mRNAs present in the cells, it is difficult to estimate accurately the proportion of HCMV RNAs that were identified in the differential display analysis. However, since we identified 2 of about 10 immediate-early mRNAs, it seems likely that the screen identified substantially less than half of the viral mRNAs that were present, even though multiple clones were isolated that corresponded to several of the viral transcripts. Partial cDNA clones corresponding to the most abundant immediate-early (IE1/IE2: ref. 38, 39) and early (TRL4: ref. 40) mRNAs were isolated, so our screen might have favored the identification of the more plentiful species.

Given the proportion of immediate-early viral mRNAs that were identified in the screen, it seems likely that we also identified substantially less than half of the cellular RNAs that were induced at 8 h after infection. Nevertheless, multiple partial cDNA clones corresponding to some of the cellular transcripts were isolated (Table 1, *supra*). In fact, 8 overlapping clones were isolated that corresponded to one of the cellular RNAs whose sequence was not found in a BLAST search.

All of the cellular RNAs that were induced at 8 h after infection proved to be interferon-inducible (Table 1 and Fig. 2C). We presume that they are induced by HCMV infection at the level of transcription as is the case when their accumulation is induced by interferon, but we have not yet determined this. A complete cDNA corresponding to one of the interferon-inducible RNAs (*cig* 49) has been cloned and sequenced. It is related to ISG54K (24). One of the partial cDNA sequences (*cig*42) also appears to be related to ISG54K, and the other 4 are not related in their primary sequence to known genes.

We were concerned that the cellular RNAs identified in the screen might be induced by interferon or another contaminant of the virus preparations, but a variety of observations argue that the induction is mediated by virus particles. The most direct evidence supporting this view derives from neutralization experiments (Fig. 3B), and the timing of the induction is not consistent with a role for interferon (Fig. 5). Further, it is unlikely that the induction involves a cytokine or small molecule other than interferon in the virus preparations since the inducing activity fractionated with the virions (Fig. 3A). We have ruled out the possibility that interferon or another signaling molecule is synthesized by infected cells and secreted to act at the cell surface, since the interferon-responsive mRNAs are induced in the presence of cycloheximide (Fig. 3A). Finally, experiments in which infected cells were mixed with uninfected cells (Fig. 4B) argue that pre-existing stores of a signaling molecule are not released after infection with HCMV to act at the cell surface and initiate a signal cascade.

A constituent of the virus particle, rather than a viral gene product synthesized after infection, mediates the induction because UV-irradiated particles that fail to express immediate-early mRNAs (Fig. 1) can sponsor the accumulation (Fig. 2A). We are currently working to identify the inducer and its mode of action.

Three different strains of HCMV strongly induced the accumulation of interferon response RNAs (Fig. 3A), and the AD169 strain was shown to induce these RNAs in HF cells prepared from three different tissue samples (data not shown). Adenovirus did not induce and HSV-1 generated a very weak induction (Fig. 3A and 5). Thus, the relatively strong HCMV-mediated induction is not a general feature of infection by DNA viruses. Adenovirus has been shown to block the induction of interferon response genes through the action of its E1A proteins (41-43). However, an E1A-deficient adenovirus mutant, *dl*312 (21), also failed to induce the genes (data not shown). In contrast, HSV-1 has been shown to induce the production of interferon-α in human peripheral mononuclear cells (44-46). So the weak induction observed in HSV-1-infected HF cells might result from a direct induction of interferon-responsive genes, from the production of double-stranded RNA which can induce the genes or from the initial induction of interferon-β with a subsequent general induction of interferon-response genes as the secreted interferon acts at the cell surface. Besides the strength of induction, the HSV-1- and HCMV-mediated reactions differ in another important respect. HCMV induces interferon-response mRNAs very early during its replication cycle in HF cells (Fig. 5), beginning about 20 h prior to the onset of viral DNA replication. In contrast, the induction observed for HSV-1 occurs later during its more rapid replication cycle (47).

Does HCMV lack the means to prevent the accumulation of interferon-inducible genes or does it somehow exploit their induction? Perhaps HCMV, in contrast to some other viruses, has not evolved the means to block the induction of interferon-inducible mRNAs. The anti-viral actions of the induced cellular products could be

antagonized by viral products at a post-transcriptional level, or HCMV might activate these genes as part of a strategy to slow and minimize the extent of its replication within an infected host. Such a strategy, together with the ability to undergo latency could facilitate the long term association of the pathogen with its host. It is also possible that the virus utilizes a component of the interferon-response pathway to activate its own genes.

References

1. Yurochko, A. D., Hwang, E-S., Rasmussen, L., Keay, S., Pereira, L. & Huang, E-S. (1997) *J. Virol.* **71**, 5051-5059.

2. Nowak, B., Gmiener, A., Sarnow, P., Levine, A. J., & Fleckenstein, B. (1984) *Virology* **134**, 91-102.

3. Roby, C. & Gibson, W. (1986) *J. Virol.* **59**, 714-727.

4. Ruger, B., Klages, S., Walla, B., Albrecht, J., Fleckenstein, B., Tomlinson, P. & Barrell, B. (1987) *J. Virol.* **61**, 446-453.

5. Lu, B. & Stinski, M. F. (1992) *J. Virol.* **66**, 4434-4444.

6. Pizzorno, M. C., OíHare, P., Sha, L., LaFemina, R. L. & Hayward, G. S. (1988) *J. Virol.* **62**, 1167-1179.

7. Malone, C. L., Vesole, D. H. & Stinski, M. F. (1990) *J. Virol.* **64**, 1498-1506.

8. Stenberg, R. M., Fortney, J., Barlow, S. W., Magrane, B. P., Nelson, J. A. & Ghazal, P. (1990) *J. Virol.* **64**, 1556-1565.

9. Klucher, K., Sommer, M., Kadonaga, J. T. & Spector, D. H. (1993) *Mol. Cell. Biol.* **13**, 1238-1250.

10. Lukac, D. M., Manuppello, J. R. & Alwine, J. C. (1994) *J. Virol.* **68**, 5184-5193.

11. Jault, F. M., Jault, J-M., Ruchti, F., Fortunato, E. A., Clark, C., Corbeil, J., Richman, D. D. & Spector, D. H. (1995) *J. Virol.* **69**, 6697-6704.

12. Bresnahan, W. A., Boldogh, I., Thompson, E. A. & Albrecht, T. (1996) *Virology* **224**, 150-160.

13. Lu, M. & Shenk, T. (1996) *J. Virol.* **70**, 8850-8857.

14. Dittmer, D. & Mocarski, E. S. (1997) *J. Virol.* **71**, 1629-1634.

15. Mocarski, E. S., Jr. (1995) in Fields Virology, Third Edition, ed. Fields, B. N., Knipe, D. M., & Howley, P. M. (Lippencott-Raven, Philadelphia), pp. 2447-2492.

16. Liang, P. & Pardee, A. B. (1992) *Science* **257**, 967-971.

17. Liang, P., Bauer, D., Averboukh, L., Warthoe, P., Rohrwild, M., Muller, H., Strauss, M. & Pardee, A. B. (1995) *Methods Enzymol.* **254**, 304-321.

18. Elek, S. D. & Stern, H. (1974) *Lancet* **1**, 1-5.

19. Plotkin, S. A., Farquhar, J. & Hornberger, E. (1976) *J. Infect. Dis.* **134**, 470-475.

20. Quinnan, G. V., Delery, M., Rook, A. H., Frederick, W. R., Epstein, J. S., Manischewitz, J. F., Jackson, L., Ramsey, K. M., Mittal, K., Plotkin, S. A. & Hilleman, M. R. (1984) *Ann. Intern. Med.* **101**, 478-483.

21. Jones, N. C. & Shenk, T. (1979) *Cell* **17**, 683-689.

22. Baldick C. J. & Shenk, T. (1996) *J. Virol.* **70**, 6097-6105.

23. Yao F. & Courtney, R. (1992) *J. Virol.* **66**: 2709-2716.

24. Levy, D., Larner, A., Chaudhuri, A., Babiss, L. E. & Darnell, J. E., Jr. (1986) *Proc. Natl. Acad. Sci., USA* **83**, 8929-8933.

25. Nomura, N., Nagase, T., Sazuka, T., Tanaka, A., Sato, S., Seki, N., Kawarabayasi, Y., Ishikawa, K. & Tabata, S. (1994) *DNA Res.* **1**, 223-229.

26. Bereta, J. & Bereta, M. (1995) *Biochem. Biophys. Res. Comm.* **217**, 363-369.

27. Cheng, Y. S., Becker-Manley, M. F., Chow, T. P. & Horan, D. C. (1985) *J. Biol. Chem.* **260**, 15834-15835.

28. Church, S. L. (1990) *Biochim. Biophys. Acta* **1087**, 250-252.

29. Kitamura, A., Takahashi, K., Okajima, A. & Kitamura, N. (1994) *Eur. J. Biochem.* **224**, 877-883.

30. Buwitt, U., Flohr, T. & Bottger, E. C. (1992) *EMBO J.* **11**, 489-496.

31. Shiojiri, S., Fukunaga, R., Ichii, Y. & Sokawa, Y. (1986) *J. Biochem.* **99**, 1455-1464.

32. Cheng, Y-S, Patterson, C. E., & Staeheli, P. (1991) *Mol. Cell. Biol.* **11**, 4717-4725.

33. Aebi, M., Fah, J., Hurt, N., Samuel, C. E., Thomis, D., Bazzigher, L., Pavlovic, J., Haller, O. & Staeheli, P. (1989) *Mol. Cell. Biol.* **9**, 5062-5072.

34. Horisberger, M. A., McMaster, G. K., Zeller, H., Wathelet, M. G., & Content, J. (1990) *J. Virol.* **64**, 1171-1181.

35. Blomstrom, D. C., Fatey, D., Kutny, R., Korant, D. & E. Knight. (1986) *J. Biol. Chem.* **261**, 8811-8816.

36. Reich, N., Evans, B., Levy, D., Fatey, D., Knight, E. & Darnell, J. E., Jr. (1987) *Proc. Natl. Acad. Sci., USA* **84**, 6394-6398.

37. Hiscott, J., Nguyen, H., & Lin, R. (1995) *Seminars Virol.* **6**, 161-173.

38. Wathen, M. W. & Stinski, M. F. (1982) *J. Virol.* **41**, 462-477.

39. McDonough, S. H. & Spector, D. H. (1983) *Virology* **125**, 31-46.

40. McDonough, S. H., Staprans, S. I. & Spector, D. H. (1985) *J. Virol.* **53**, 711-718.

41. Reich, N., Pine, R., Levy, D. & Darnell, J. E., Jr. (1988) *J. Virol.* **62**, 114-119.

42. Gutch, M. J. & Reich, N. C. (1991) *Proc. Natl. Acad. Sci., USA* **88**, 7913-7917.

43. Kalvakolanu, D. V. R., Bandyopadhyay, S. K., Harter, M. L. & Sen, G. C. (1991) *Proc. Natl. Acad. Sci., USA* **88**, 7459-7463.

44. Fitzgerald, P. A., Von Wussow, P. & Lopez, C. (1982) *J. Immunol.* **129**, 819-824.

45. Feldman, M. & Fitzgerald-Bocarsly, P. (1990) *J. Interferon Res.* **10**, 435-446.

46. Li, Q., Feldman, M., Harmon, C. & Fitzgerald-Bocarsly, P. (1996) *J. Interferon and Cytokine Res.* **1**, 109-118.

47. Roizman, B. & Sears, A. E. (1995) in Fields Virology, Third Edition, ed Fields, B.N., Knipe, D.M., & Howley, P.M. (Lippencott-Raven, Philadelphia), pp. 2231-2295.

While the invention has been described and illustrated herein by references to various specific material, procedures and examples, it is understood that the invention is not restricted to the particular material combinations of material, and procedures selected for that purpose. Numerous variations of such details can be implied as will be appreciated by those skilled in the art.

Various references are cited throughout this Specification, each of which is incorporated herein by reference in its entirety.

WHAT IS CLAIMED IS:

1. A set of human genes, the expression of which, is specifically modulated by human cytamegalovirus (HCMV) and limited to the following:

a) genes that are induced to express by both HCMV and interferon, designated HCMV-inducible genes (*cig* or *cigs*); and,

b) genes that repressed in the presence of HCMV infection, designated HCMV-repressible genes (*crg* or *crgs*).

2. A *cig* of Claim 1 which is a cDNA having a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, and 32.

3. A *cig* of Claim 1 which is a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 27, 29, 31, and 33.

4. A *crg* of Claim 1 which is a cDNA having a nucleotide sequence selected from the group consisting of SEQ ID NOS:34, 36, 38, and 39.

5. A *crg* of Claim 1 which is a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID 35, and 37.

6. A DNA sequence that hybridizes to any of the nucleotide sequences of Claim 2 or 4, and degenerate varients thereof.

7. A recombinant DNA molecule comprising a DNA sequence of Claim 2 or 4, and degenerate variants thereof.

8. The recombinant DNA molecule of either of Claim 7, wherein said DNA sequence is operatively linked to an expression control sequence.

9. The recombinant DNA molecule of Claim 8, wherein said expression control sequence is selected from the group consisting of the early or late promoters of SV40 or adenovirus, the *lac* system, the *trp* system, the *TAC* system, the *TRC* system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase, the promoters of acid phosphatase and the promoters of the yeast α-mating factors.

10. A probe capable of screening for the *cigs* or *crgs* in alternate species prepared from the DNA sequence of Claim 6.

11. A unicellular host transformed with a recombinant DNA molecule comprising a DNA sequence or degenerate variant thereof, which encodes a *cig* or *crg* gene product, or a fragment thereof, selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21-26, 28, 30, 32, 34, 36, 38, and 39, wherein said DNA sequence is operatively linked to an expression control sequence.

12. The unicellular host of Claim 11 wherein the unicellular host is selected from the group consisting of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, yeasts, CHO, R1.1, B-W, L-M, COS 1, COS 7, BSC1, BSC40, and BMT10 cells, plant cells, insect cells, and human cells in tissue culture.

13. A method for detecting the level of expression of *cig* or *crg* mRNAs consisting of:

A. capture probes, based on the sequences of Claim 6, immobilized onto a solid support;

B. contacting a biological sample containing *cig* and *crg* mRNAs from a human or human cell culture with the capture probes under standard hybridization conditions; and,

C. detecting the levels of hybridization that has occured between the target mRNAs and the capture probe;

wherein the levels of hybridization detected reveals the levels of expression from the *cigs* and *crgs* of Claim 1.

14. The method of Claim 13 used as a screening assay to identify drugs or compounds that alter the expression of *cig* or *crg* mRNAs, and are thus candidates for anti-viral or anti-HCMV drugs.

15. The method of Claim 13 used as a diagnostic assay to evaluate the efficacy of a treatment regimen for HCMV or other viral infections.

16. An antibody to a polypeptide sequence of Claim 3 or 5.

17. The antibody of Claim 16 which is a polyclonal antibody.

18. The antibody of Claim 16 which is a monoclonal antibody.

19. An immortal cell line that produces a monoclonal antibody according to Claim 18.

20. The antibody of Claim 16 labeled with a detectable label.

21. The antibody of Claim 20 wherein the label is selected from enzymes, chemicals which fluoresce and radioactive elements.

22. An antisense nucleic acid against a *cig* mRNA comprising a nucleic acid sequence hybridizing to said mRNA.

23. The antisense nucleic acid of Claim 22 which is RNA.

24. The antisense nucleic acid of Claim 22 which is DNA.

25. The antisense nucleic acid of Claim 22 which binds to the initiation codon of any of said mRNAs.

26. A recombinant DNA molecule having a DNA sequence which, on transcription, produces an antisense ribonucleic acid against a *cig* mRNA, said antisense ribonucleic acid comprising an nucleic acid sequence capable of hybridizing to said mRNA.

27. A *cig* gene product-producing cell line transfected with the recombinant DNA molecule of Claim 26.

28. A method for creating a cell line which exhibits reduced expression of a cig mRNA, comprising transfecting a cig mRNA-producing cell line with a recombinant DNA molecule of claim 26.

29. A ribozyme that cleaves cig mRNA.

30. The ribozyme of Claim 29 which is a *Tetrahymena*-type ribozyme.

31. The ribozyme of Claim 29 which is a Hammerhead-type ribozyme.

32. A recombinant DNA molecule having a DNA sequence which, upon transcription, produces the ribozyme of claim 29.

33. A *cig* mRNA-producing cell line transfected with the recombinant DNA molecule of claim 32.

34. A method for creating a cell line which exhibits reduced expression of a *cig* mRNA, comprising transfecting a cig mRNA-producing cell line with the recombinant DNA molecule of claim 29.

35. A *crg* gene product (protein) used as a n anti-viral or anti-HCMV therapeutic.

36. A *cig* gene product (protein) used in conjunction with interferon therapy to reduce toxicity of said interferon and thus allow administration of higher doses of said interferon.

FIG.1A

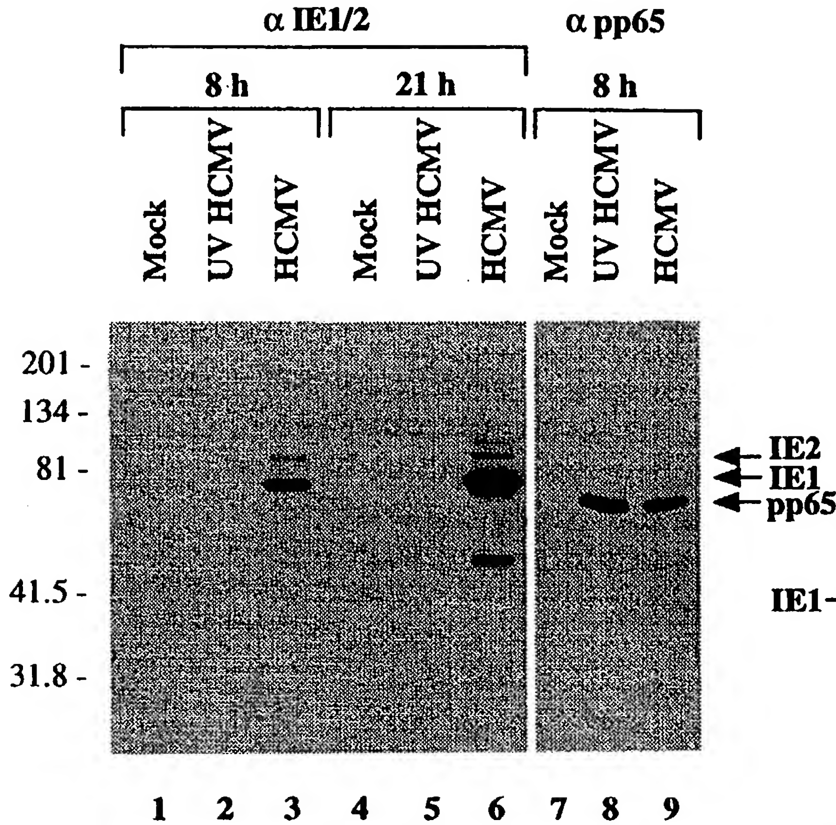
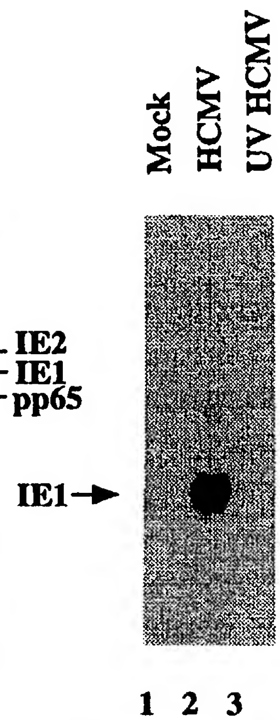


FIG.1B



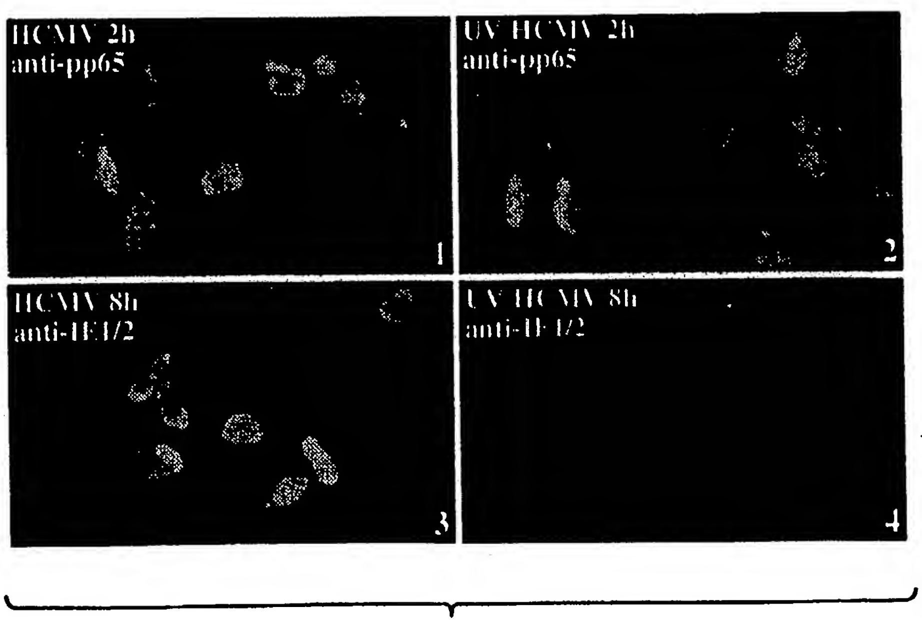


FIG.1C

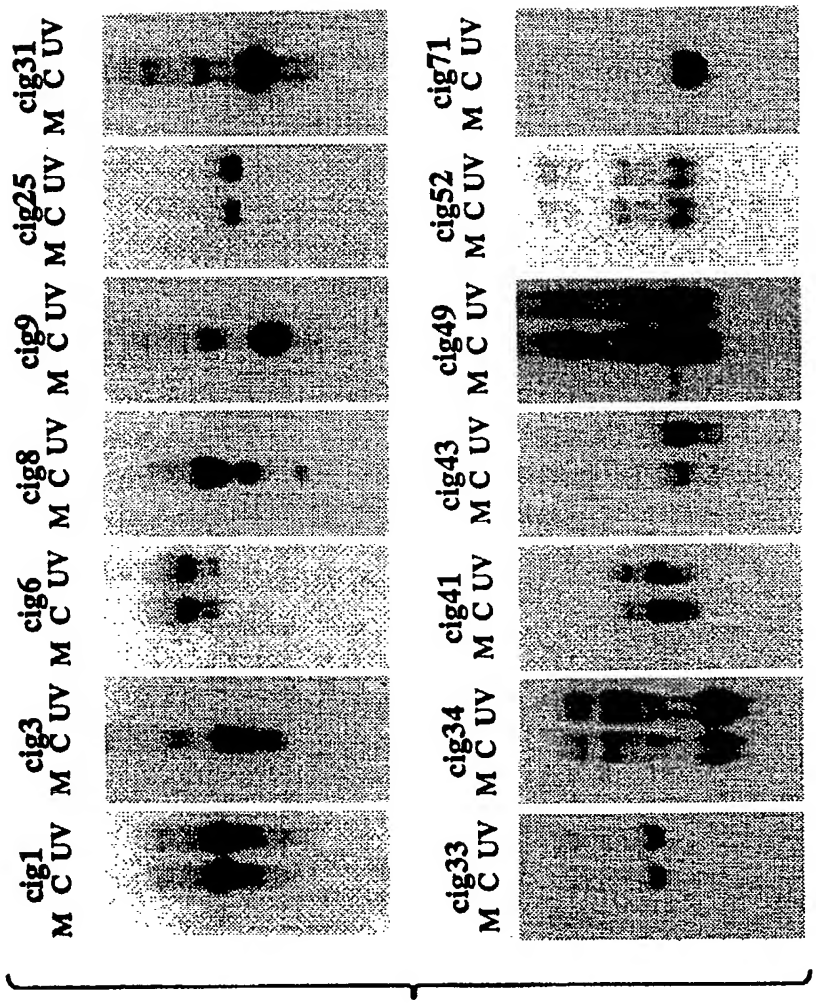

cig1 M C UV
cig3 M C UV
cig6 M C UV
cig8 M C UV
cig9 M C UV
cig25 M C UV
cig31 M C UV
cig33 M C UV
cig34 M C UV
cig41 M C UV
cig43 M C UV
cig49 M C UV
cig52 M C UV
cig71 M C UV

FIG.2A

FIG.2B

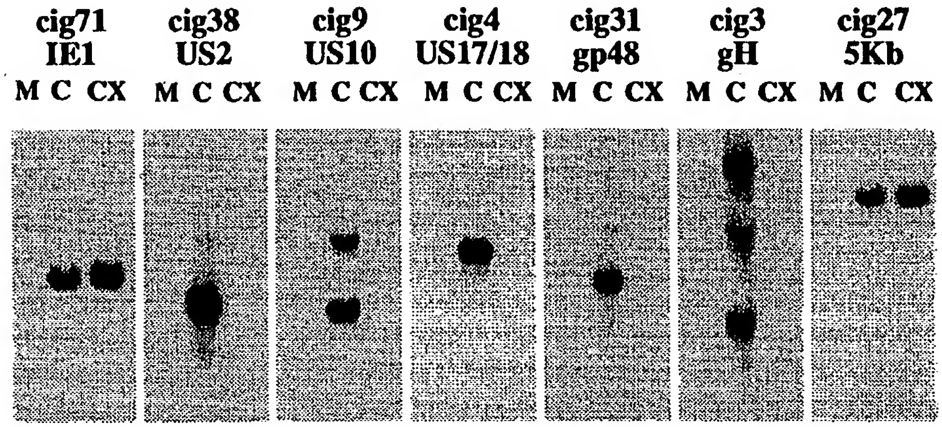


FIG.2C

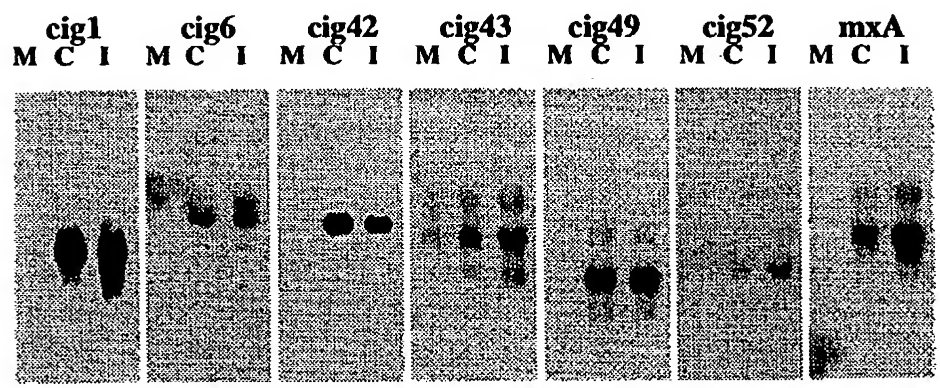


FIG.2D

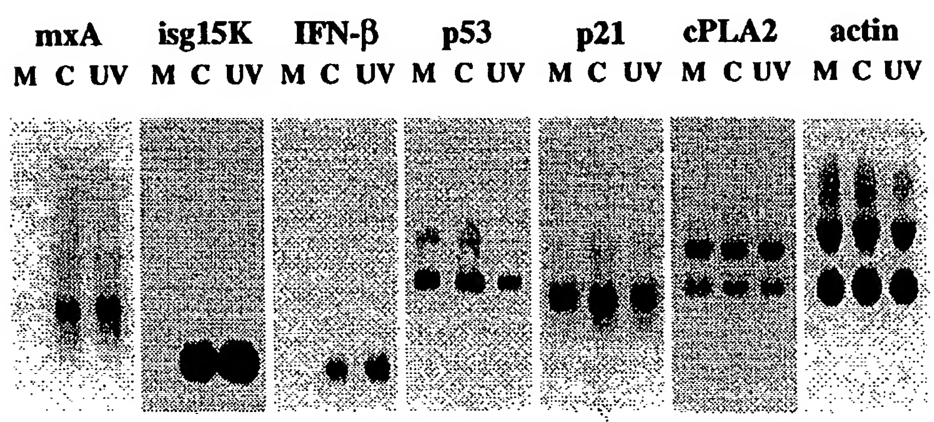


FIG.3A

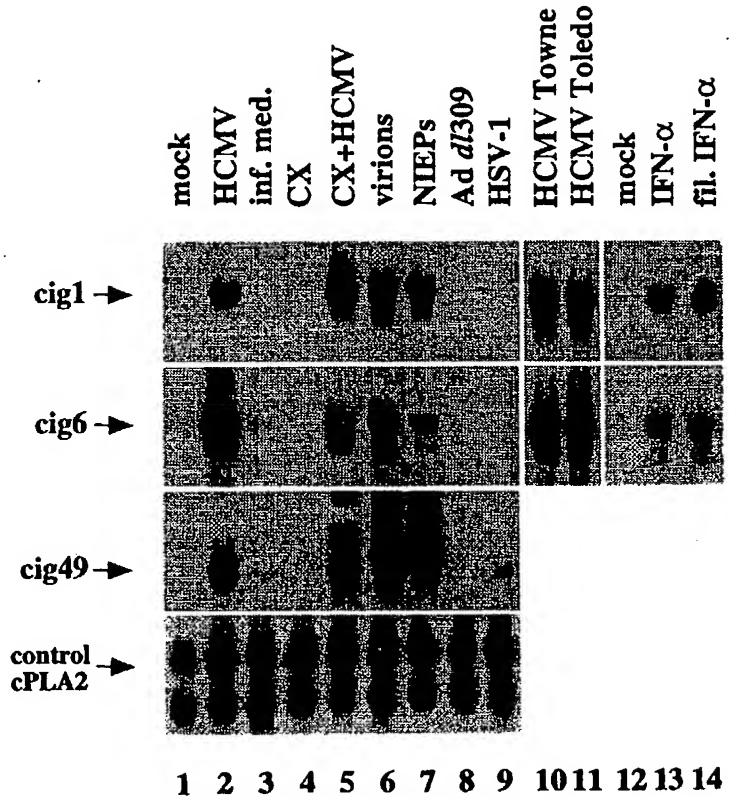

mock
HCMV
inf. med.
CX
CX+HCMV
virions
NIEPs
Ad dl309
HSV-1
HCMV Towne
HCMV Toledo
mock
IFN-α
fil. IFN-α
cig1
cig6
cig49
control cPLA2
1 2 3 4 5 6 7 8 9 10 11 12 13 14

FIG.3B

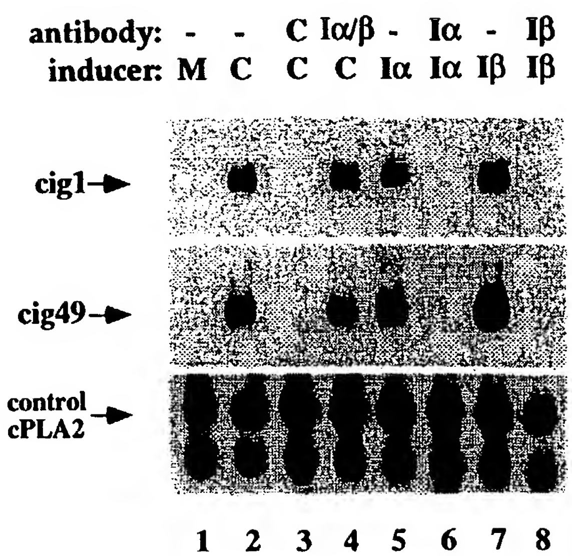

antibody: - - C Iα/β - Iα - Iβ
inducer: M C C C Iα Iα Iβ Iβ
cig1
cig49
control cPLA2
1 2 3 4 5 6 7 8

FIG.4A

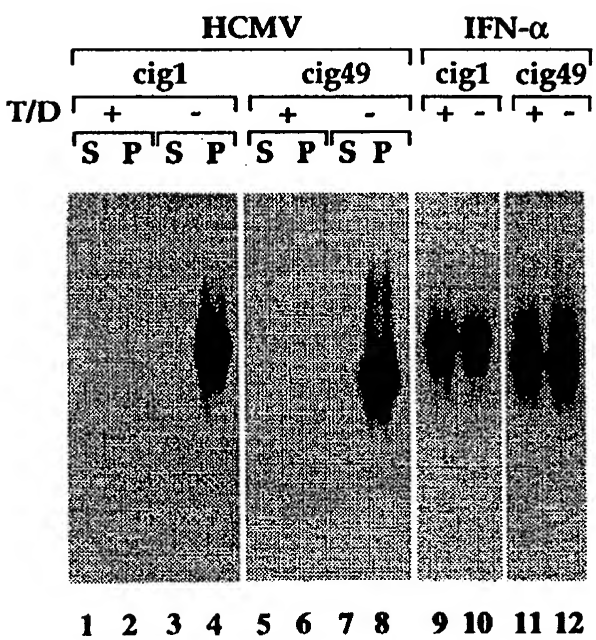

HCMV
IFN-α
cig1
cig49
cig1
cig49
T/D
+
-
+
-
+
-
+
-
S P S P S P S P
1 2 3 4 5 6 7 8 9 10 11 12

FIG.4B

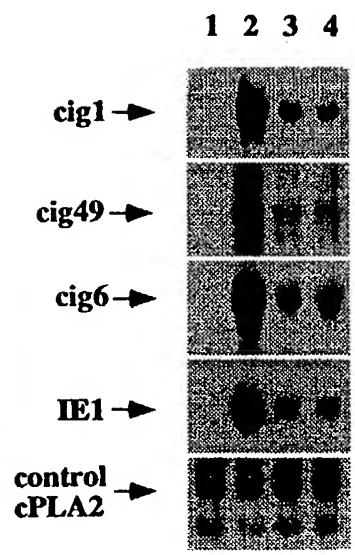
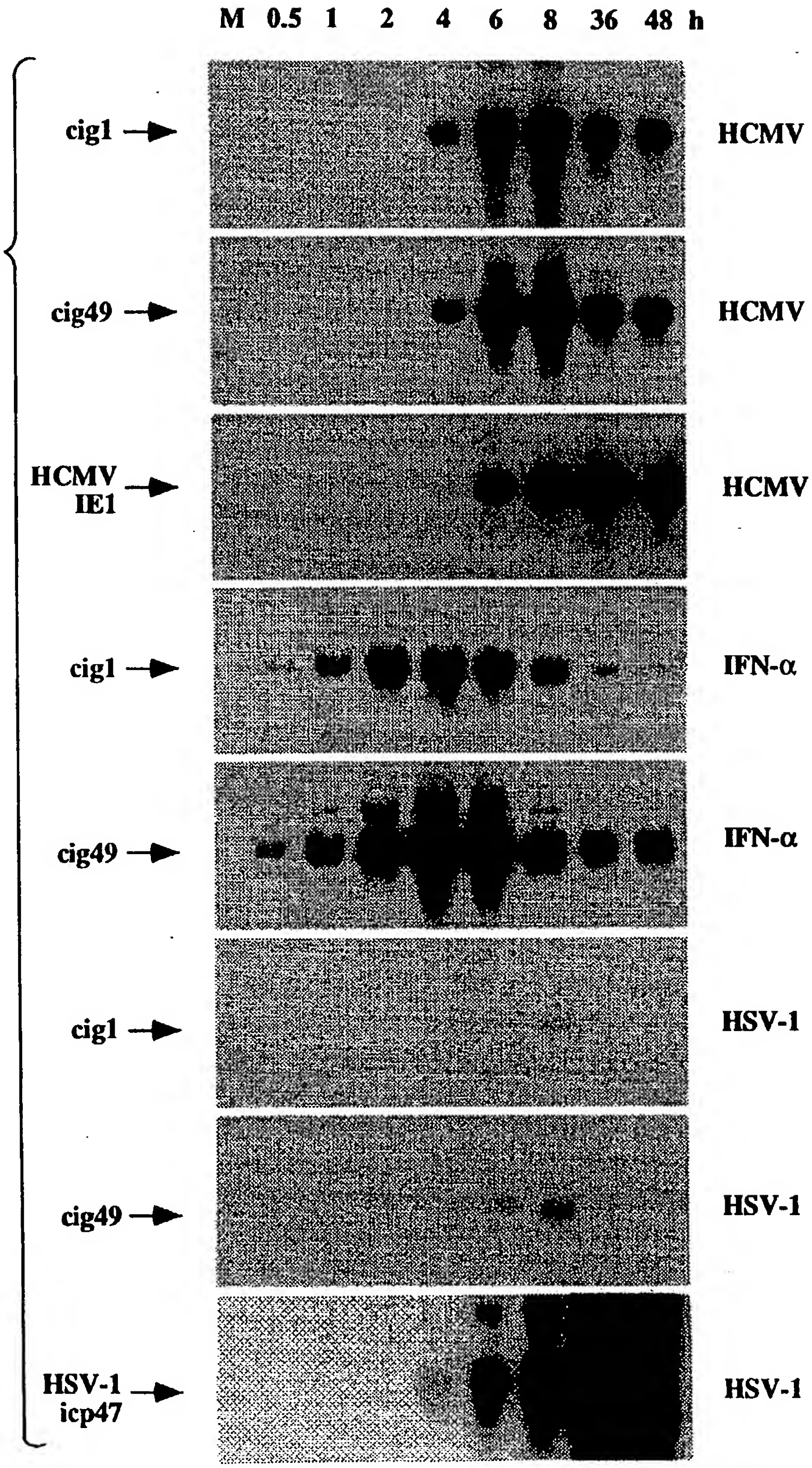

1 2 3 4
cig1
cig49
cig6
IE1
control
cPLA2

FIG.5


M 0.5 1 2 4 6 8 36 48 h
cig1
HCMV
cig49
HCMV
HCMV IE1
HCMV
cig1
IFN-α
cig49
IFN-α
cig1
HSV-1
cig49
HSV-1
HSV-1 icp47
HSV-1

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Zhu, Hua
Cong, Jiang-Ping
Schenk, Thomas

(ii) TITLE OF INVENTION: HUMAN GENES REGULATED BY HUMAN CYTOMEGALOVIURS AND INTERFERON

(iii) NUMBER OF SEQUENCES: 39

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: David A. Jackson, Esq.

(B) STREET: 411 Hackensack Ave, Continental Plaza, 4th Floor

(C) CITY: Hackensack

(D) STATE: New Jersey

(E) COUNTRY: USA

(F) ZIP: 07601

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Jackson Esq., David A.

(B) REGISTRATION NUMBER: 26,742

(C) REFERENCE/DOCKET NUMBER: 2275-1-001 P1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 201-487-5800

(B) TELEFAX: 201-343-1684

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 280 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TATTAACCCT CACAAAATGT GGTGGACCAA AGTCTAATAG GGCTCAGTAT CCCCCATCGC 60

TTATCTCTGC CTCCTTCCTC CTCTTCCCAG TCTATCATCA ACCTTGAGTA TTTACACAAT 120

GTGAATTCAA GTGCCTGATT AATTGAGGTG GCAACATAGT TTGAGACGAG GGCAGAGAAC 180

AGGAAGATAC ATAGCTAGAA GCGACGGGTA CAAAAAGCAA TGTGTACAAG AAGACTTTCA 240

GCAAGTATAC AGAGAGTTCA CCTCTACTCT GCCCTCCTCA

280

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Leu Thr Leu Thr Lys Cys Gly Gly Pro Lys Ser Asn Arg Ala Gln Tyr
1 5 10 15

Pro Pro Ser Leu Ile Ser Ala Ser Phe Leu Leu Phe Pro Val Tyr His
20 25 30

Gln Pro

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5378 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGATCCCCTG CTGGGAGGGG GCAGGGGACC TGTTCCCACC GTGTGCCCAA GACCTCTTTT
60

CCCACTTTTT CCCTCTTCTT GACTCACCCT GCCCTCAATA TCCCCCGGCG CAGCAGTGAA
120

AGGGAGTCCC TGGCTCCTGG CTCGCCTGCA CGTCCCAGGG CGGGGAGGGA CTTCCGCCCT
180

CACGTCCCGC TCTTCGCCCC AGGCTGGATG GAATGAAAGG CACACTGTCT CTCTCCCTAG
240

GCAGCACAGC CCACAGGTTT CAGGAGTGCC TTTGTGGGAG GCCTCTGGGC CCCCACCAGC
300

CATCCTGTCC TCCGCCTGGG GCCCCAGCCC GGAGAGAGCC GCTGGTGCAC ACAGGGCCGG
360

GATTGTCTGC CCTAATTATC AGGTCCAGGC TACAGGGCTG CAGGACATCG TGACCTTCCG
420

TGCAGAAACC TCCCCCTCCC CCTCAAGCCG CCTCCCGAGC CTCCTTCCTC TCCAGGCCCC
480

CAGTGCCCAG TGCCCAGTGC CCAGCCCAGG CCTCGGTCCC AGAGATGCCA GGAGCCAGGA
540

GATGGGGAGG GGGAAGTGGG GGCTGGGAAG GAACCACGGG CCCCCGCCCG AGCCCATGGG
600

CCCCTCCTAG GCCTTTGCCT GAGCAGACCG GTGTCACTAC CGCAGAGCCT CGAGGAGAAG
660

TTCCCCAACT TTCCCGCCTC TCAGCCTTTG AAAGAAAGAA AGGGGAGGGG GCAGGCCGCG
720

TGCAGCCGCG AGCGGTGCTG GGCTCCGGCT CCAATTCCCC ATCTCAGTCG TTCCCAAAGT
780

CCTCCTGTTT CATCCAAGCG TGTAAGGGTC CCCGTCCTTG ACTCCCTAGT GTCCTGCTGC
840

CCACAGTCCA GTCCTGGGAA CCAGCACCGA TCACCTCCCA TCGGGCCAAT CTCAGTCCCT
900

TCCCCCTACG TCGGGGCCCA CACGCTCGGT GCGTGCCCAG TTGAACCAGG CGGCTGCGGA
960

AAAAAAAAAG CGGGGAGAAA GTAGGGCCCG GCTACTAGCG GTTTTACGGG CGCACGTAGC
1020

TCAGGCCTCA AGACCTTGGG CTGGGACTGG CTGAGCCTGG CGGGAGGCGG GGTCCGAGTC
1080

ACCGCCTGCC GCCGCGCCCC CGGTTTCTAT AAATTGAGCC CGCAGCCTCC CGCTTCGCTC
1140

TCTGCTCCTC CTGTTCGACA GTCAGCCGCA TCTTCTTTTG CGTCGCCAGG TGAAGACGGG
1200

CGGAGAGAAA CCCGGGAGGC TAGGGACGGC CTGAAGGCGG CAGGGGCGGG CGCAGGCCGG
1260

ATGTGTTCGC GCCGCTGCGG GGTGGGCCCG GGCGGCCTCC GCATTGCAGG GGCGGGCGGA
1320

GGACGTGATG CGGCGCGGGC TGGGCATGGA GGCCTGGTGG GGGAGGGGAG GGGAGGCGTG
1380

TGTGTCGGCC GGGGCCACTA GGCGCTCACT GTTCTCTCCC TCCGCGCAGC CGAGCCACAT
1440

CGCTCAGACA CCATGGGGAA GGTGAAGGTC GGAGTCAACG GGTGAGTTCG CGGGTGGCTG
1500

GGGGGCCCTG GGCTGCGACC GCCCCCGAAC CGCGTCTACG AGCCTTGCGG GCTCCGGGTC
1560

TTTGCAGTCG TATGGGGGCA GGGTAGCTGT TCCCCGCAAG GAGAGCTCAA GGTCAGCGCT
1620

CGGACCTGGC GGAGCCCCGC ACCCAGGCTG TGGCGCCCTG TGCAGCTCCG CCCTTGCGGC
1680

GCCATCTGCC CGGAGCCTCC TTCCCCTAGT CCCCAGAAAC AGGAGGTCCC TACTCCCGCC
1740

CGAGATCCCG ACCCGGACCC CTAGGTGGGG GACGCTTTCT TTCCTTTCGC GCTCTGCGGG
1800

GTCACGTGTC GCAGAGGAGC CCCTCCCCCA CGGCCTCCGG CACCGCAGGC CCCGGGATGC
1860

TAGTGCGCAG CGGGTGCATC CCTGTCCGGA TGCTGCGCCT GCGGTAGAGC GGCCGCCATG
1920

TTGCAACCGG GAAGGAAATG AATGGGCAGC CGTTAGGAAA GCCTGCCGGT GACTAACCCT
1980

GCGCTCCTGC CTCGATGGGT GGAGTCGCGT GTGGCGGGGA AGTCAGGTGG AGCGAGGCTA
2040

GCTGGCCCGA TTTCTCCTCC GGGTGATGCT TTTCCTAGAT TATTCTCTGG TAAATCAAAG
2100

AAGTGGGTTT ATGGAGGTCC TCTTGTGTCC CCTCCCCGCA GAGGTGTGGT GGCTGTGGCA
2160

TGGTGCCAAG CCGGGAGAAG CTGAGTCATG GGTAGTTGGA AAAGGACATT TCCACCGCAA
2220

AATGGCCCCT CTGGTGGTGG CCCCTTCCTG CAGCGGCTCA CCTCACGGCC CCGCCCTTCC
2280

CCTGCCAGCC TAGCGTTGAC CCGACCCCAA AGGCCAGGCT GTAAATGTCA CCGGGAGGAT
2340

TGGGTGTCTG GGCGCCTCGG GGAACCTGCC CTTCTCCCCA TTCCGTCTTC CGGAAACCAG
2400

ATCTCCACCG CACCCTGGTC TGAGGTCTGA GGTTAAATAT AGCTGCTGAC CTTTCTGTAG
2460

CTGGGGGCCT GGGCTGGGGC TCTCTCCCAT CCCTTCTCCC CACACACATG CACTTACCTG
2520

TGCTCCCACT CCTGATTTCT GGAAAAGAGC TAGGAAGGAC AGGCAACTTG GCAAATCAAA
2580

GCCCTGGGAC TAGGGGGTTA AAATACAGCT TCCCCTCTTC CCACCCGCCC CAGTCTCTGT
2640

CCCTTTTGTA GGAGGGACTT AGAGAAGGGG TGGGCTTGCC CTGTCCAGTT AATTTCTGAC
2700

CTTTACTCCT GCCCTTTGAG TTTGATGATG CTGAGTGTAC AAGCGTTTTC TCCCTAAAGG
2760

GTGCAGCTGA GCTAGGCAGC AGCAAGCATT CCTGGGGTGG CATAGTGGGG TGGTGAATAC
2820

CATGTACAAA GCTTGTGCCC AGACTGTGGG TGGCAGTGCC CACATGGCCG CTTCTCCTGG
2880

AAGGGCTTCG TATGACTGGG GGTGTTGGGC AGCCCTGGAG CCTTCAGTTG CAGCCATGCC
2940

TTAAGCCAGG CCAGCCTGGC AGGGAAGCTC AAGGGAGATA AAATTCAACC TCTTGGGCCC
3000

TCCTGGGGGT AAGGAGATGC TGCATTCGCC CTCTTAATGG GGAGGTGGCC TAGGGCTGCT
3060

CACATATTCT GGAGGAGCCT CCCCTCCTCA TGCCTTCTTG CCTCTTGTCT CTTAGATTTG
3120

GTCGTATTGG GCGCCTGGTC ACCAGGGCTG CTTTTAACTC TGGTAAAGTG GATATTGTTG
3180

CCATCAATGA CCCCTTCATT GACCTCAACT ACATGGTGAG TGCTACATGG TGAGCCCCAA
3240

AGCTGGTGTG GGAGGAGCCA CCTGGCTGAT GGGCAGCCCC TTCATACCCT CACGTATTCC
3300

CCCAGGTTTA CATGTTCCAA TATGATTCCA CCCATGGCAA ATTCCATGGC ACCGTCAAGG
3360

CTGAGAACGG GAAGCTTGTC ATCAATGGAA ATCCCATCAC CATCTTCCAG GAGTGAGTGG
3420

AAGACAGAAT GGAAGAAATG TGCTTTGGGG AGGCAACTAG GATGGTGTGG CTCCCTTGGG
3480

TATATGGTAA CCTTGTGTCC CTCAATATGG TCCTGTCCCC ATCTCCCCCC CACCCCGGTA
3540

GGCGAGATCC CTCCAAAATC AAGTGGGGCG ATGCTGGCGC TGAGTACGTC GTGGAGTCCA
3600

CTGGCGTCTT CACCACCATG GAGAAGGCTG GGGTGAGTGC AGGAGGGCCC GCGGGAGGGG
3660

AAGCTGACTC AGCCCTGCAA AGGCAGGACC CGGGTTCATA ACTGTCTGCT TCTCTGCTGT
3720

AGGCTCATTT GCAGGGGGGA GCCAAAAGGG TCATCATCTC TGCCCCCTCT GCTGATGCCC
3780

CCATGTTCGT CATGGGTGTG AACCATGAGA AGTATGACAA CAGCCTCAAG ATCATCAGGT
3840

GAGGAAGGCA GGGCCCGTGG AGAAGCGGCC AGCCTGGCAC CCTATGGACA CGCTCCCCTG
3900

ACTTGCGCCC CGCTCCCTCT TTCTTTGCAG CAATGCCTCC TGCACCACCA ACTGCTTAGC
3960

ACCCCTGGCC AAGGTCATCC ATGACAACTT TGGTATCGTG GAAGGACTCA TGGTATGAGA
4020

GCTGGGGAAT GGGACTGAGG CTCCCACCTT TCTCATCCAA GACTGGCTCC TCCCTGCTGG
4080

GGCTGCGTGC AACCCTGGGG TTGGGGGTTC TGGGGACTGG CTTTCCCATA ATTTCCTTTC
4140

AAGGTGGGGA GGGAGGTAGA GGGGTGATGT GGGGAGTACG CTGCAGGGCC TCACTCCTTT
4200

TGCAGACCAC AGTCCATGCC ATCACTGCCA CCCAGAAGAC TGTGGATGGC CCCTCCGGGA
4260

AACTGTGGCG TGATGGCCGC GGGGCTCTCC AGAACATCAT CCCTGCCTCT ACTGGCGCTG
4320

CCAAGGCTGT GGGCAAGGTC ATCCCTGAGC TGAACGGGAA GCTCACTGGC ATGGCCTTCC
4380

GTGTCCCCAC TGCCAACGTG TCAGTGGTGG ACCTGACCTG CCGTCTAGAA AAACCTGCCA
4440

AATATGATGA CATCAAGAAG GTGGTGAAGC AGGCGTCGGA GGGCCCCCTC AAGGGCATCC
4500

TGGGCTACAC TGAGCACCAG GTGGTCTCCT CTGACTTCAA CAGCGACACC CACTCCTCCA
4560

CCTTTGACGC TGGGGCTGGC ATTGCCCTCA ACGACCACTT TGTCAAGCTC ATTTCCTGGT
4620

ATGTGGCTGG GGCCAGAGAC TGGCTCTTAA AAAGTGCAGG GTCTGGCGCC CTCTGGTGGC
4680

TGGCTCAGAA AAAGGGCCCT GACAACTCTT TTCATCTTCT AGGTATGACA ACGAATTTGG
4740

CTACAGCAAC AGGGTGGTGG ACCTCATGGC CCACATGGCC TCCAAGGAGT AAGACCCCTG
4800

GACCACCAGC CCCAGCAAGA GCACAAGAGG AAGAGAGAGA CCCTCACTGC TGGGGAGTCC
4860

CTGCCACACT CAGTCCCCCA CCACACTGAA TCTCCCCTCC TCACAGTTGC CATGTAGACC
4920

CCTTGAAGAG GGGAGGGGCC TAGGGAGCCG CACCTTGTCA TGTACCATCA ATAAAGTACC
4980

CTGTGCTCAA CCAGTTACTT GTCCTGTCTT ATTCTAGGGT CTGGGGCAGA GGGGAGGGAA
5040

GCTGGGCTTG TGTCAAGGTG AGACATTCTT GCTGGGGAGG GACCTGGTAT GTTCTCCTCA
5100

GACTGAGGGT AGGGCCTCCA AACAGCCTTG CTTGCTTCGA GAACCATTTG CTTCCCGCTC
5160

AGACGTCTTG AGTGCTACAG GAAGCTGGCA CCACTACTTC AGAGAACAAG GCCTTTTCCT
5220

CTCCTCGCTC CAGTCCTAGG CTATCTGCTG TTGGCCAAAC ATGGAAGAAG CTATTCTGTG
5280

GGCAGCCCCA GGGAGGCTGA CAGGTGGAGG AAGTCAGGGC TCGCACTGGG CTCTGACGCT
5340

GACTGGTTAG TGGAGCTCAG CCTGGAGCTG AGCTGCAG
5378

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 335 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Gly Lys Val Lys Val Gly Val Asn Gly Phe Gly Arg Ile Gly Arg
1               5                   10                  15

Leu Val Thr Arg Ala Ala Phe Asn Ser Gly Lys Val Asp Ile Val Ala
            20                  25                  30

Ile Asn Asp Pro Phe Ile Asp Leu Asn Tyr Met Val Tyr Met Phe Gln
        35                  40                  45

Tyr Asp Ser Thr His Gly Lys Phe His Gly Thr Val Lys Ala Glu Asn
    50                  55                  60

Gly Lys Leu Val Ile Asn Gly Asn Pro Ile Thr Ile Phe Gln Glu Arg
65                  70                  75                  80

Asp Pro Ser Lys Ile Lys Trp Gly Asp Ala Gly Ala Glu Tyr Val Val
                85                  90                  95

Glu Ser Thr Gly Val Phe Thr Thr Met Glu Lys Ala Gly Ala His Leu
            100                 105                 110

Gln Gly Gly Ala Lys Arg Val Ile Ile Ser Ala Pro Ser Ala Asp Ala
        115                 120                 125

Pro Met Phe Val Met Gly Val Asn His Glu Lys Tyr Asp Asn Ser Leu
    130                 135                 140
```

```
Lys Ile Ile Ser Asn Ala Ser Cys Thr Thr Asn Cys Leu Ala Pro Leu
145                 150                 155                 160

Ala Lys Val Ile His Asp Asn Phe Gly Ile Val Glu Gly Leu Met Thr
                165                 170                 175

Thr Val His Ala Ile Thr Ala Thr Gln Lys Thr Val Asp Gly Pro Ser
            180                 185                 190

Gly Lys Leu Trp Arg Asp Gly Arg Gly Ala Leu Gln Asn Ile Ile Pro
        195                 200                 205

Ala Ser Thr Gly Ala Ala Lys Ala Val Gly Lys Val Ile Pro Glu Leu
    210                 215                 220

Asn Gly Lys Leu Thr Gly Met Ala Phe Arg Val Pro Thr Ala Asn Val
225                 230                 235                 240

Ser Val Val Asp Leu Thr Cys Arg Leu Glu Lys Pro Ala Lys Tyr Asp
                245                 250                 255

Asp Ile Lys Lys Val Val Lys Gln Ala Ser Glu Gly Pro Leu Lys Gly
            260                 265                 270

Ile Leu Gly Tyr Thr Glu His Gln Val Val Ser Ser Asp Phe Asn Ser
        275                 280                 285

Asp Thr His Ser Ser Thr Phe Asp Ala Gly Ala Gly Ile Ala Leu Asn
    290                 295                 300

Asp His Phe Val Lys Leu Ile Ser Trp Tyr Asp Asn Glu Phe Gly Tyr
305                 310                 315                 320

Ser Asn Arg Val Val Asp Leu Met Ala His Met Ala Ser Lys Glu
                325                 330                 335
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2881 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGAAGTGC TAGAAGCCAG TGCTCGTGAA CTAAGGAGAA AAAGAACAGA CAAGGGAACA
60

GCCTGGACAT GGCATCAGAG ATCCACATGA CAGGCCCAAT GTGCCTCATT GAGAACACTA
120

ATGGGCGACT GATGGCGAAT CCAGAAGCTC TGAAGATCCT TTCTGCCATT ACACAGCCTA
180

TGGTGGTGGT GGCAATTGTG GGCCTCTACC GCACAGGCAA ATCCTACCTG ATGAACAAGC
240

TGGCTGGAAA GAAAAAGGGC TTCTCTCTGG GCTCCACGGT GCAGTCTCAC ACTAAAGGAA
300

TCTGGATGTG GTGTGTGCCC CACCCCAAGA AGCCAGGCCA CATCCTAGTT CTGCTGGACA
360

CCGAGGGTCT GGGAGATGTA GAGAAGGGTG ACAACCAGAA TGACTCCTGG ATCTTCGCCC
420

TGGCCGTCCT CCTGAGCAGC ACCTTCGTGT ACAATAGCAT AGGAACCATC AACCAGCAGG
480

CTATGGACCA ACTGTACTAT GTGACAGAGC TGACACATAG AATCCGATCA AAATCCTCAC
540

CTGATGAGAA TGAGAATGAG GTTGAGGATT CAGCTGACTT TGTGAGCTTC TTCCCAGACT
600

TTGTGTGGAC ACTGAGAGAT TTCTCCCTGG ACTTGGAAGC AGATGGACAA CCCCTCACAC
660

CAGATGAGTA CCTGACATAC TCCCTGAAGC TGAAGAAAGG TACCAGTCAA AAAGATGAAA
720

CTTTTAACCT GCCCAGACTC TGTATCCGGA AATTCTTCCC AAAGAAAAAA TGCTTTGTCT
780

TTGATCGGCC CGTTCACCGC AGGAAGCTTG CCCAGCTCGA GAAACTACAA GATGAAGAGC
840

TGGACCCCGA ATTTGTGCAA CAAGTAGCAG ACTTCTGTTC CTACATCTTT AGTAATTCCA
900

AAACTAAAAC TCTTTCAGGA GGCATCCAGG TCAACGGGCC TCGTCTAGAG AGCCTGGTGC
960

TGACCTACGT CAATGCCATC AGCAGTGGGG ATCTGCCGTG CATGGAGAAC GCAGTCCTGG
1020

CCTTGGCCCA GATAGAGAAC TCAGCTGCAG TGCAAAAGGC TATTGCCCAC TATGAACAGC
1080

AGATGGGCCA GAAGGTGCAG CTGCCCACAG AAAGCCTCCA GGAGCTGCTG GACCTGCACA
1140

GGGACAGTGA GAGAGAGGCC ATTGAAGTCT TCATCAGGAG TTCCTTCAAA GATGTGGACC
1200

ATCTATTTCA AAAGGAGTTA GCGGCCCAGC TAGAAAAAAA GCGGGATGAC TTTTGTAAAC
1260

AGAATCAGGA AGCATCATCA GATCGTTGCT CAGGTTTACT TCAGGTCATT TTCAGTCCTC
1320

TAGAAGAAGA AGTGAAGGCG GGAATTTATT CGAAACCAGG GGGCTATCGT CTCTTTGTTC
1380

AGAAGCTACA AGACCTGAAG AAAAAGTACT ATGAGGAACC GAGGAAGGGG ATACAGGCTG
1440

AAGAGATTCT GCAGACATAC TTGAAATCCA AGGAGTCTAT GACTGATGCA ATTCTCCAGA
1500

CAGACCAGAC TCTCACAGAA AAAGAAAAGG AGATTGAAGT GGAACGTGTG AAAGCTGAGT
1560

CTGCACAGGC TTCAGCAAAA ATGTTGCAGG AAATGCAAAG AAAGAATGAG CAGATGATGG
1620

AACAGAAGGA GAGGAGTTAT CAGGAACACT TGAAACAACT GACTGAGAAG ATGGAGAACG
1680

ACAGGGTCCA GTTGCTGAAA GAGCAAGAGA GGACCCTCGC TCTTAAACTT CAGGAACAGG
1740

AGCAACTACT AAAAGAGGGA TTTCAAAAAG AAAGCAGAAT AATGAAAAAT GAGATACAGG
1800

ATCTCCAGAC GAAAATGAGA CGACGAAAGG CATGTACCAT AAGCTAAAGA CCAGAGCCTT
1860

CCTGTCACCC CTAACCAAGG CATAATTGAA ACAATTTTAG AATTTGGAAC AAGCGTCACT 1920

ACATTTGATA ATAATTAGAT CTTGCATCAT AACACCAAAA GTTTATAAAG GCATGTGGTA 1980

CAATGATCAA AATCATGTTT TTTCTTAAAA AAAAAAAAAA GACTGTAAAT TGTGCAACAA 2040

AGATGCATTT ACCTCTGTAT CAACTCAGGA AATCTCATAA GCTGGTACCA CTCAGGAGAA 2100

GTTTATTCTT CCAGATGACC AGCAGTAGAC AAATGGATAC TGAGCAGAGT CTTAGGTAAA 2160

AGTCTTGGGA AATATTTGGG CATTGGTCTG GCCAAGTCTA CAATGTCCCA ATATCAAGGA 2220

CAACCACCCT AGCTTCTTAG TGAAGACAAT GTACAGTTAT CCATTAGATC AAGACTACAC 2280

GGTCTATGAG CAATAATGTG ATTTCTGGAC ATTGCCCATG TATAATCCTC ACTGATGATT 2340

TCAAGCTAAA GCAAACCACC TTATACAGAG ATCTAGAATC TCTTTATGTT CTCCAGAGGA 2400

AGGTGGAAGA AACCATGGGC AGGAGTAGGA ATTGAGTGAT AAACAATTGG GCTAATGAAG 2460

AAAACTTCTC TTATTGTTCA GTTCATCCAG ATTATAACTT CAATGGGACA CTTTAGACCA 2520

TTAGACAATT GACACTGGAT TAAACAAATT CACATAATGC CAAATACACA ATGTATTTAT 2580

AGCAACGTAT AATTTGCAAA GATGGACTTT AAAAGATGCT GTGTAACTAA ACTGAAATAA
2640

TTCAATTACT TATTATTTAG AATGTTAAAG CTTATGATAG TCTTTTCTAA TTCTTAACAC
2700

TCATACTTGA AATCTTTCCG AGTTTCCCCA GAAGAGAATA TGGGATTTTT TTTGACATTT
2760

TTGACCCATT TAATAATGCT CTTGTGTTTA CCTAGTATAT GTAGACTTTG TCTTATGTGT
2820

CAAAAGTCCT AGGAAAGTGG TTGATGTTTC TTATAGCAAT TAAAAATTAT TTTTGAACTG
2880

A
2881

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 592 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Ser Glu Ile His Met Thr Gly Pro Met Cys Leu Ile Glu Asn
1 5 10 15

Thr Asn Gly Arg Leu Met Ala Asn Pro Glu Ala Leu Lys Ile Leu Ser
20 25 30

Ala Ile Thr Gln Pro Met Val Val Val Ala Ile Val Gly Leu Tyr Arg
35 40 45

Thr Gly Lys Ser Tyr Leu Met Asn Lys Leu Ala Gly Lys Lys Lys Gly
50 55 60

Phe Ser Leu Gly Ser Thr Val Gln Ser His Thr Lys Gly Ile Trp Met
65 70 75 80

Trp Cys Val Pro His Pro Lys Lys Pro Gly His Ile Leu Val Leu Leu
85 90 95

Asp Thr Glu Gly Leu Gly Asp Val Glu Lys Gly Asp Asn Gln Asn Asp
100 105 110

Ser Trp Ile Phe Ala Leu Ala Val Leu Leu Ser Ser Thr Phe Val Tyr
115 120 125

Asn Ser Ile Gly Thr Ile Asn Gln Gln Ala Met Asp Gln Leu Tyr Tyr
130 135 140

Val Thr Glu Leu Thr His Arg Ile Arg Ser Lys Ser Ser Pro Asp Glu
145 150 155 160

Asn Glu Asn Glu Val Glu Asp Ser Ala Asp Phe Val Ser Phe Phe Pro
165 170 175

Asp Phe Val Trp Thr Leu Arg Asp Phe Ser Leu Asp Leu Glu Ala Asp
180 185 190

Gly Gln Pro Leu Thr Pro Asp Glu Tyr Leu Thr Tyr Ser Leu Lys Leu

195 200 205

Lys Lys Gly Thr Ser Gln Lys Asp Glu Thr Phe Asn Leu Pro Arg Leu
210 215 220

Cys Ile Arg Lys Phe Phe Pro Lys Lys Lys Cys Phe Val Phe Asp Arg
225 230 235 240

Pro Val His Arg Arg Lys Leu Ala Gln Leu Glu Lys Leu Gln Asp Glu
245 250 255

Glu Leu Asp Pro Glu Phe Val Gln Gln Val Ala Asp Phe Cys Ser Tyr
260 265 270

Ile Phe Ser Asn Ser Lys Thr Lys Thr Leu Ser Gly Gly Ile Gln Val
275 280 285

Asn Gly Pro Arg Leu Glu Ser Leu Val Leu Thr Tyr Val Asn Ala Ile
290 295 300

Ser Ser Gly Asp Leu Pro Cys Met Glu Asn Ala Val Leu Ala Leu Ala
305 310 315 320

Gln Ile Glu Asn Ser Ala Ala Val Gln Lys Ala Ile Ala His Tyr Glu
325 330 335

Gln Gln Met Gly Gln Lys Val Gln Leu Pro Thr Glu Ser Leu Gln Glu
340 345 350

Leu Leu Asp Leu His Arg Asp Ser Glu Arg Glu Ala Ile Glu Val Phe
355 360 365

Ile Arg Ser Ser Phe Lys Asp Val Asp His Leu Phe Gln Lys Glu Leu
370 375 380

Ala Ala Gln Leu Glu Lys Lys Arg Asp Asp Phe Cys Lys Gln Asn Gln
385 390 395 400

Glu Ala Ser Ser Asp Arg Cys Ser Gly Leu Leu Gln Val Ile Phe Ser
405 410 415

Pro Leu Glu Glu Glu Val Lys Ala Gly Ile Tyr Ser Lys Pro Gly Gly
420 425 430

Tyr Arg Leu Phe Val Gln Lys Leu Gln Asp Leu Lys Lys Lys Tyr Tyr
435 440 445

Glu Glu Pro Arg Lys Gly Ile Gln Ala Glu Glu Ile Leu Gln Thr Tyr
450 455 460

Leu Lys Ser Lys Glu Ser Met Thr Asp Ala Ile Leu Gln Thr Asp Gln
465 470 475 480

Thr Leu Thr Glu Lys Glu Lys Glu Ile Glu Val Glu Arg Val Lys Ala
485 490 495

Glu Ser Ala Gln Ala Ser Ala Lys Met Leu Gln Glu Met Gln Arg Lys
500 505 510

Asn Glu Gln Met Met Glu Gln Lys Glu Arg Ser Tyr Gln Glu His Leu
515 520 525

Lys Gln Leu Thr Glu Lys Met Glu Asn Asp Arg Val Gln Leu Leu Lys
530 535 540

Glu Gln Glu Arg Thr Leu Ala Leu Lys Leu Gln Glu Gln Glu Gln Leu
545 550 555 560

Leu Lys Glu Gly Phe Gln Lys Glu Ser Arg Ile Met Lys Asn Glu Ile
565 570 575

Gln Asp Leu Gln Thr Lys Met Arg Arg Arg Lys Ala Cys Thr Ile Ser
580 585 590

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 976 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCGGGCGGCG CAGGAGCGGC ACTCGTGGCT GTGGTGGCTT CGGCAGCGGC TTCAGCAGAT 60

CGGCGGCATC AGCGGTAGCA CCAGCACTAG CAGCATGTTG AGCCGGGCAG TGTGCGGCAC 120

CAGCAGGCAG CTGGCTCCGG CTTTGGGGTA TCTGGGCTCC AGGCAGAAGC ACAGCCTCCC 180

CGACCTGCCC TACGACTACG GCGCCCTGGA ACCTCACATC AACGCGCAGA TCATGCAGCT 240

GCACCACAGC AAGCACCACG CGGCCTACGT GAACAACCTG AACGTCACCG AGGAGAAGTA 300

CCAGGAGGCG TTGGCCAAGG GAGATGTTAC AGCCCAGACA GCTCTTCAGC CTGCACTGAA 360

GTTCAATGGT GGTGGTCATA TCAATCATAG CATTTTCTGG ACAAACCTCA GCCCTAACGG
420

TGGTGGAGAA CCCAAAGGGG AGTTGCTGGA AGCCATCAAA CGTGACTTTG GTTCCTTTGA
480

CAAGTTTAAG GAGAAGCTGA CGGCTGCATC TGTTGGTGTC CAAGGCTCAG GTTGGGGTTG
540

GCTTGGTTTC AATAAGGAAC GGGGACACTT ACAAATTGCT GCTTGTCCAA ATCAGGATCC
600

ACTGCAAGGA ACAACAGGCC TTATTCCACT GCTGGGGATT GATGTGTGGG AGCACGCTTA
660

CTACCTTCAG TATAAAAATG TCAGGCCTGA TTATCTAAAA GCTATTTGGA ATGTAATCAA
720

CTGGGAGAAT GTAACTGAAA GATACATGGC TTGCAAAAAG TAAACCACGA TCGTTATGCT
780

GAGTATGTTA AGCTCTTTAT GACTGTTTTT GTAGTGGTAT AGAGTACTGC AGAATACAGT
840

AAGCTGCTCT ATTGTAGCAT TTCTTGATGT TGCTTAGTCA CTTATTTCAT AAACAACTTA
900

ATGTTCTGAA TAATTTCTTA CTAAACATTT TGTTATTGGG CAAGTGATTG AAAATAGTAA
960

ATGCTTTGTG TGATTG
976

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 211 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Leu Ser Arg Ala Val Cys Gly Thr Ser Arg Gln Leu Ala Ala Leu
1   5   10   15

Gly Tyr Leu Gly Ser Arg Gln Lys His Ser Leu Asp Leu Tyr Asp Tyr
            20   25   30

Gly Ala Leu Glu His Ile Asn Ala Gln Ile Met Gln Leu His His Ser
        35   40   45

Lys His His Ala Ala Tyr Val Asn Asn Leu Asn Val Thr Glu Glu Lys
    50   55   60

Tyr Gln Glu Ala Leu Ala Lys Gly Asp Val Thr Ala Gln Thr Ala Leu
65   70   75   80

Gln Ala Leu Lys Phe Asn Gly Gly Gly His Ile Asn His Ser Ile Phe
                85   90   95

Trp Thr Asn Leu Ser Asn Gly Gly Gly Glu Lys Gly Glu Leu Leu Glu
            100   105   110

Ala Ile Lys Arg Asp Phe Gly Ser Phe Asp Lys Phe Lys Glu Lys Leu

115 120 125

Thr Ala Ala Ser Val Gly Val Gln Gly Ser Gly Trp Gly Trp Leu Gly
130 135 140

Phe Asn Lys Glu Arg Gly His Leu Gln Ile Ala Ala Cys Asn Gln Asp
145 150 155 160

Leu Gln Gly Thr Thr Gly Leu Ile Leu Leu Gly Ile Asp Val Trp Glu
165 170 175

His Ala Tyr Tyr Leu Gln Tyr Lys Asn Val Arg Asp Tyr Leu Lys Ala
180 185 190

Ile Trp Asn Val Ile Asn Trp Glu Asn Val Thr Glu Arg Tyr Met Ala
195 200 205

Cys Lys Lys
210

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1335 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGCAGTGA CAACTCGTTT GACATGGTTG CACGAAAAGA TCCTGCAAAA TCATTTTGGA 60

GGGAAGCGGC TTAGCCTTCT CTATAAGGGT AGTGTCCATG GATTCCGTAA TGGAGTTTTG 120

CTTGACAGAT GTTGTAATCA AGGGCCTACT CTAACAGTGA TTTATAGTGA AGATCATATT 180

ATTGGAGCAT ATGCAGAAGA GAGTTACCAG GAAGGAAAGT ATGCTTCCAT CATCCTTTTT 240

GCACTTCAAG ATACTAAAAT TTCAGAATGG AAACTAGGAC TATGTACACC AGAAACACTG 300

TTTTGTTGTG ATGTTACAAA ATATAACTCC CCAACTAATT TCCAGATAGA TGGAAGAAAT 360

AGAAAAGTGA TTATGGACTT AAAGACAATG GAAAATCTTG GACTTGCTCA AAATTGTACT 420

ATCTCTATTC AGGATTATGA AGTTTTTCGA TGCGAAGATT CACTGGATGA AAGAAAGATA 480

AAAGGGGTCA TTGAGCTCAG GAAGAGCTTA CTGTCTGCCT TGAGAACTTA TGAACCATAT 540

GGATCCCTGG TTCAACAAAT ACGAATTCTC CTCCTGGGTC CAATTGGAGC TCCCAAGTCC 600

AGCTTTTTCA ACTCAGTGAG GTCTGTTTTC CAAGGGCATG TAACGCATCA GGCTTTGGTG 660

GGCACTAATA CAACTGGGAT ATCTGAGAAG TATAGGACAT ACTCTATTAG AGACGGGAAA 720

GATGGCAAAT ACCTGCCGTT TATTCTGTGT GACTCACTGG GGCTGAGTGA GAAAGAAGGC
780

GGCCTGTGCA GGGATGACAT ATTCTATATC TTGAACGGTA ACATTCGTGA TAGATACCAG
840

TTTAATCCCA TGGAATCAAT CAAATTAAAT CATCATGACT ACATTGATTC CCCATCGCTG
900

AAGGACAGAA TTCATTGTGT GGCATTTGTA TTTGATGCCA GCTCTATTCA ATACTTCTCC
960

TCTCAGATGA TAGTAAAGAT CAAAAGAATT CAAAGGGAGT TGGTAAACGC TGGTGTGGTA
1020

CATGTGGCTT TGCTCACTCA TGTGGATAGC ATGGATTTGA TTACAAAAGG TGACCTTATA
1080

GAAATAGAGA GATGTGAGCC TGTGAGGTCC AAGCTAGAGG AAGTCCAAAG AAAACTTGGA
1140

TTTGCTCTTT CTGACATCTC GGTGGTTAGC AATTATTCCT CTGAGTGGGA GCTGGACCCT
1200

GTAAAGGATG TTCTAATTCT TTCTGCTCTG AGACGAATGC TATGGGCTGC AGATGACTTC
1260

TTAGAGGATT TGCCTTTTGA GCAAATAGGG AATCTAAGGG AGGAAATTAT CAACTGTGCA
1320

CAAGGAAAAA AATAG
1335

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 444 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Ala Val Thr Thr Arg Leu Thr Trp Leu His Glu Lys Ile Leu Gln
1               5                   10                  15

Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly Ser Val
            20                  25                  30

His Gly Phe Arg Asn Gly Val Leu Leu Asp Arg Cys Cys Asn Gln Gly
        35                  40                  45

Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile Ile Gly Ala Tyr
    50                  55                  60

Ala Glu Glu Ser Tyr Gln Glu Gly Lys Tyr Ala Ser Ile Ile Leu Phe
65                  70                  75                  80

Ala Leu Gln Asp Thr Lys Ile Ser Glu Trp Lys Leu Gly Leu Cys Thr
                85                  90                  95

Pro Glu Thr Leu Phe Cys Cys Asp Val Thr Lys Tyr Asn Ser Pro Thr
            100                 105                 110

Asn Phe Gln Ile Asp Gly Arg Asn Arg Lys Val Ile Met Asp Leu Lys
```

115 120 125

Thr Met Glu Asn Leu Gly Leu Ala Gln Asn Cys Thr Ile Ser Ile Gln
130 135 140

Asp Tyr Glu Val Phe Arg Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile
145 150 155 160

Lys Gly Val Ile Glu Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr
165 170 175

Tyr Glu Pro Tyr Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu Leu
180 185 190

Gly Pro Ile Gly Ala Pro Lys Ser Ser Phe Phe Asn Ser Val Arg Ser
195 200 205

Val Phe Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr
210 215 220

Thr Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys
225 230 235 240

Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu Ser
245 250 255

Glu Lys Glu Gly Gly Leu Cys Arg Asp Asp Ile Phe Tyr Ile Leu Asn
260 265 270

Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu Ser Ile Lys
275 280 285

Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu Lys Asp Arg Ile
290 295 300

His Cys Val Ala Phe Val Phe Asp Ala Ser Ser Ile Gln Tyr Phe Ser
305 310 315 320

```
Ser Gln Met Ile Val Lys Ile Lys Arg Ile Gln Arg Glu Leu Val Asn
                325                 330                 335

Ala Gly Val Val His Val Ala Leu Leu Thr His Val Asp Ser Met Asp
            340                 345                 350

Leu Ile Thr Lys Gly Asp Leu Ile Glu Ile Glu Arg Cys Glu Pro Val
        355                 360                 365

Arg Ser Lys Leu Glu Glu Val Gln Arg Lys Leu Gly Phe Ala Leu Ser
    370                 375                 380

Asp Ile Ser Val Val Ser Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro
385                 390                 395                 400

Val Lys Asp Val Leu Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala
                405                 410                 415

Ala Asp Asp Phe Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu
            420                 425                 430

Arg Glu Glu Ile Ile Asn Cys Ala Gln Gly Lys Lys
        435                 440
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2567 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTGCTGAACA AATCCTCTGA CCTCAGGCCG GCTGTGAACG TAGTTCCTGA GAGATAGCAA 60

ACATGCCCAA CAGTGAGCCC GCATCTCTGC TGGAGCTGTT CAACAGCATC GCCACACAAG 120

GGGAGCTCGT AAGGTCCCTC AAAGCGGGAA ATGCGTCAAA GGATGAAATT GATTCTGCAG 180

TAAAGATGTT GGTGTCATTA AAAATGAGCT ACAAAGCTGC CGCGGGGGAG GATTACAAGG 240

CTGACTGTCC TCCAGGGAAC CCAGCACCTA CCAGTAATCA TGGCCCAGAT GCCACAGAAG 300

CTGAAGAGGA TTTTGTGGAC CCATGGACAG TACAGACAAG CAGTGCAAAA GGCATAGACT 360

ACGATAAGCT CATTGTTCGG TTTGGAAGTA GTAAAATTGA CAAAGAGCTA ATAAACCGAA 420

TAGAGAGAGC CACCGGCCAA AGACCACACC ACTTCCTGCG CAGAGGCATC TTCTTCTCAC 480

ACAGAGATAT GAATCAGGTT CTTGATGCCT ATGAAAATAA GAAGCCATTT TATCTGTACA 540

CGGGCCGGGG CCCCTCTTCT GAAGCAATGC ATGTAGGTCA CCTCATTCCA TTTATTTTCA 600

CAAAGTGGCT CCAGGATGTA TTTAACGTGC CCTTGGTCAT CCAGATGACG GATGACGAGA
660

AGTATCTGTG GAAGGACCTG ACCCTGGACC AGGCCTATAG CTATGCTGTG GAGAATGCCA
720

AGGACATCAT CGCCTGTGGC TTTGACATCA ACAAGACTTT CATATTCTCT GACCTGGACT
780

ACATGGGGAT GAGCTCAGGT TTCTACAAAA ATGTGGTGAA GATTCAAAAG CATGTTACCT
840

TCAACCAAGT GAAAGGCATT TTCGGCTTCA CTGACAGCGA CTGCATTGGG AAGATCAGTT
900

TTCCTGCCAT CCAGGCTGCT CCCTCCTTCA GCAACTCATT CCCACAGATC TTCCGAGACA
960

GGACGGATAT CCAGTGCCTT ATCCCATGTG CCATTGACCA GGATCCTTAC TTTAGAATGA
1020

CAAGGGACGT CGCCCCCAGG ATCGGCTATC CTAAACCAGC CCTGTTGCAC TCCACCTTCT
1080

TCCCAGCCCT GCAGGGCGCC CAGACCAAAA TGAGTGCCAG CGACCCCAAC TCCTCCATCT
1140

TCCTCACCGA CACGGCCAAG CAGATCAAAA CCAAGGTCAA TAAGCATGCG TTTTCTGGAG
1200

GGAGAGACAC CATCGAGGAG CACAGGCAGT TTGGGGGCAA CTGTGATGTG GACGTGTCTT
1260

TCATGTACCT GACCTTCTTC CTCGAGGACG ACGACAAGCT CGAGCAGATC AGGAAGGATT
1320

ACACCAGCGG ACGCATGCTC ACCGGTGAGC TCAAGAAGGC ACTCATAGAG GTTCTGCAGC
1380

CCTTGATCGC AGAGCACCAG GCCCGGCGCA AGGAGGTCAC GGATGAGATA GTGAAAGAGT
1440

TCATGACTCC CCGGAAGCTG TCCTTCGACT TTCAGTAGCA CTCGTTTTAC ATATGCTTAT
1500

AAAAGAAGTG ATGTATCAGT AATGTATCAA TAATCCCAGC CCAGTCAAAG CACCGCCACC
1560

TGTAGGCTTC TGTCTCATGG TAATTACTGG GCCTGGCCTC TGTAAGCCTG TGTATGTTAT
1620

CAATACTGTT TCTTCCTGTG AGTTCCATTA TTTCTATCTC TTATGGGCAA AGCATTGTGG
1680

GTAATTGGTG CTGGCTAACA TTGCATGGTC GGATAGAGAA GTCCAGCTGT GAGTCTCTCC
1740

CCAAAGCAGC CCCACAGTGG AGCCTTTGGC TGGAAGTCCA TGGGCCACCC TGTTCTTGTC
1800

CATGGAGGAC TCCGAGGGTT CCAAGTATAC TCTTAAGACC CACTCTGTTT AAAAATATAT
1860

ATTCTATGTA TGCGTATATG GAATTGAAAT GTCATTATTG TAACCTAGAA AGTGCTTTGA
1920

AATATTGATG TGGGGAGGTT TATTGAGCAC AAGATGTATT TCAGCCCATG CCCCCTCCCA
1980

AAAAGAAATT GATAAGTAAA AGCTTCGTTA TACATTTGAC TAAGAAATCA CCCAGCTTTA
2040

AAGCTGCTTT TAACAATGAA GATTGAACAG AGTTCAGCAA TTTTGATTAA ATTAAGACTT
2100

GGGGGTGAAA CTTTCCAGTT TACTGAACTC CAGACCATGC ATGTAGTCCA CTCCAGAAAT
2160

CATGCTCGCT TCCCTTGGCA CACCAGTGTT CTCCTGCCAA ATGACCCTAG ACCCTCTGTC
2220

CTGCAGAGTC AGGGTGGCTT TTCCCCTGAC TGTGTCCGAT GCCAAGGAGT CCTGGCCTCC
2280

GCAGATGCTT CATTTTGACC CTTGGCTGCA GTGGAAGTCA GCACAGAGCA GTGCCCTGGC
2340

TGTGTCCTGG ACGGGTGGAC TTAGCTAGGG AGAAAGTCGA GGCAGCAGCC CTCGAGGCCC
2400

TCACAGATGT CTAGGCAGGC CTCATTTCAT CACGCAGCAT GTGCAGGCCT GGAAGAGCAA
2460

AGCCAAATCT CAGGGAAGTC CTTGGTTGAT GTATCTGGGT CTCCTCTGGA GCACTCTGCC
2520

CTCCTGTCAC CCAGTAGAGT AAATAAACTT CCTTGGCTCC TAAAAAA
2567

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 471 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Pro Asn Ser Glu Pro Ala Ser Leu Leu Glu Leu Phe Asn Ser Ile
1   5   10   15

Ala Thr Gln Gly Glu Leu Val Arg Ser Leu Lys Ala Gly Asn Ala Ser
            20   25   30

Lys Asp Glu Ile Asp Ser Ala Val Lys Met Leu Val Ser Leu Lys Met
        35   40   45

Ser Tyr Lys Ala Ala Ala Gly Glu Asp Tyr Lys Ala Asp Cys Pro Pro
    50   55   60

Gly Asn Pro Ala Pro Thr Ser Asn His Gly Pro Asp Ala Thr Glu Ala
65   70   75   80

Glu Glu Asp Phe Val Asp Pro Trp Thr Val Gln Thr Ser Ser Ala Lys
                85   90   95

Gly Ile Asp Tyr Asp Lys Leu Ile Val Arg Phe Gly Ser Ser Lys Ile
            100   105   110

Asp Lys Glu Leu Ile Asn Arg Ile Glu Arg Ala Thr Gly Gln Arg Pro
        115   120   125

His His Phe Leu Arg Arg Gly Ile Phe Phe Ser His Arg Asp Met Asn
    130   135   140

Gln Val Leu Asp Ala Tyr Glu Asn Lys Lys Pro Phe Tyr Leu Tyr Thr

145 150 155 160

Gly Arg Gly Pro Ser Ser Glu Ala Met His Val Gly His Leu Ile Pro
165 170 175

Phe Ile Phe Thr Lys Trp Leu Gln Asp Val Phe Asn Val Pro Leu Val
180 185 190

Ile Gln Met Thr Asp Asp Glu Lys Tyr Leu Trp Lys Asp Leu Thr Leu
195 200 205

Asp Gln Ala Tyr Ser Tyr Ala Val Glu Asn Ala Lys Asp Ile Ile Ala
210 215 220

Cys Gly Phe Asp Ile Asn Lys Thr Phe Ile Phe Ser Asp Leu Asp Tyr
225 230 235 240

Met Gly Met Ser Ser Gly Phe Tyr Lys Asn Val Val Lys Ile Gln Lys
245 250 255

His Val Thr Phe Asn Gln Val Lys Gly Ile Phe Gly Phe Thr Asp Ser
260 265 270

Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile Gln Ala Ala Pro Ser
275 280 285

Phe Ser Asn Ser Phe Pro Gln Ile Phe Arg Asp Arg Thr Asp Ile Gln
290 295 300

Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro Tyr Phe Arg Met Thr
305 310 315 320

Arg Asp Val Ala Pro Arg Ile Gly Tyr Pro Lys Pro Ala Leu Leu His
325 330 335

Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln Thr Lys Met Ser Ala
340 345 350

Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp Thr Ala Lys Gln Ile
355 360 365

Lys Thr Lys Val Asn Lys His Ala Phe Ser Gly Gly Arg Asp Thr Ile
370 375 380

Glu Glu His Arg Gln Phe Gly Gly Asn Cys Asp Val Asp Val Ser Phe
385 390 395 400

Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Asp Lys Leu Glu Gln Ile
405 410 415

Arg Lys Asp Tyr Thr Ser Gly Arg Met Leu Thr Gly Glu Leu Lys Lys
420 425 430

Ala Leu Ile Glu Val Leu Gln Pro Leu Ile Ala Glu His Gln Ala Arg
435 440 445

Arg Lys Glu Val Thr Asp Glu Ile Val Lys Glu Phe Met Thr Pro Arg
450 455 460

Lys Leu Ser Phe Asp Phe Gln
465 470

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1347 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGGGGAAGAG ATAAAAGCAA ACAGGTCTGG GAGGCAGTTC TGTTGCCACT CTCTCTCCTG 60

TCAATGATGG ATCTCAGAAA TACCCCAGCC AAATCTCTGG ACAAGTTCAT TGAAGACTAT 120

CTCTTGCCAG ACACGTGTTT CCGCATGCAA ATCAACCATG CCATTGACAT CATCTGTGGG 180

TTCCTGAAGG AAAGGTGCTT CCGAGGTAGC TCCTACCCTG TGTGTGTGTC CAAGGTGGTA 240

AAGGGTGGCT CCTCAGGCAA GGGCACCACC CTCAGAGGCC GATCTGACGC TGACCTGGTT 300

GTCTTCCTCA GTCCTCTCAC CACTTTTCAG GATCAGTTAA ATCGCCGGGG AGAGTTCATC 360

CAGGAAATTA GGAGACAGCT GGAAGCCTGT CAAAGAGAGA GAGCATTTTC CGTGAAGTTT 420

GAGGTCCAGG CTCCACGCTG GGGCAACCCC CGTGCGCTCA GCTTCGTACT GAGTTCGCTC 480

CAGCTCGGGG AGGGGGTGGA GTTCGATGTG CTGCCTGCCT TTGATGCCCT GGGTCAGTTG 540

ACTGGCAGCT ATAAACCTAA CCCCCAAATC TATGTCAAGC TCATCGAGGA GTGCACCGAC 600

CTGCAGAAAG AGGGCGAGTT CTCCACCTGC TTCACAGAAC TACAGAGAGA CTTCCTGAAG
660

CAGCGCCCCA CCAAGCTCAA GAGCCTCATC CGCCTAGTCA AGCACTGGTA CCAAAATTGT
720

AAGAAGAAGC TTGGGAAGCT GCCACCTCAG TATGCCCTGG AGCTCCTGAC GGTCTATGCT
780

TGGGAGCGAG GGAGCATGAA AACACATTTC AACACAGCCC AGGGATTTCG GACGGTCTTG
840

GAATTAGTCA TAAACTACCA GCAACTCTGC ATCTACTGGA CAAAGTATTA TGACTTTAAA
900

AACCCCATTA TTGAAAAGTA CCTGAGAAGG CAGCTCACGA AACCCAGGCC TGTGATCCTG
960

GACCCGGCGG ACCCTACAGG AAACTTGGGT GGTGGAGACC CAAAGGGTTG GAGGCAGCTG
1020

GCACAAGAGG CTGAGGCCTG GCTGAATTAC CCATGCTTTA AGAATTGGGA TGGGTCCCCA
1080

GTGAGCTCCT GGATTCTGCT GGTGAGACCT CCTGCTTCCT CCCTGCCATT CATCCCTGCC
1140

CCTCTCCATG AAGCTTGAGA CATATAGCTG GAGACCATTC TTTCCAAAGA ACTTACCTCT
1200

TGCCAAAGGC CATTTATATT CATATAGTGA CAGGCTGTGC TCCATATTTT ACAGTCATTT
1260

TGGTCACAAT CGAGGGTTTC TGGAATTTTC ACATCCCTTG TCCAGAATTC ATTCCCCTAA
1320

GAGTAATAAT AAATAATCTC TAACACC

1347

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 364 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Met Asp Leu Arg Asn Thr Pro Ala Lys Ser Leu Asp Lys Phe Ile
1 5 10 15

Glu Asp Tyr Leu Leu Pro Asp Thr Cys Phe Arg Met Gln Ile Asn His
20 25 30

Ala Ile Asp Ile Ile Cys Gly Phe Leu Lys Glu Arg Cys Phe Arg Gly
35 40 45

Ser Ser Tyr Pro Val Cys Val Ser Lys Val Val Lys Gly Gly Ser Ser
50 55 60

Gly Lys Gly Thr Thr Leu Arg Gly Arg Ser Asp Ala Asp Leu Val Val
65 70 75 80

Phe Leu Ser Pro Leu Thr Thr Phe Gln Asp Gln Leu Asn Arg Arg Gly
85 90 95

Glu Phe Ile Gln Glu Ile Arg Arg Gln Leu Glu Ala Cys Gln Arg Glu
100 105 110

Arg Ala Phe Ser Val Lys Phe Glu Val Gln Ala Pro Arg Trp Gly Asn
115 120 125

Pro Arg Ala Leu Ser Phe Val Leu Ser Ser Leu Gln Leu Gly Glu Gly
130 135 140

Val Glu Phe Asp Val Leu Pro Ala Phe Asp Ala Leu Gly Gln Leu Thr
145 150 155 160

Gly Ser Tyr Lys Pro Asn Pro Gln Ile Tyr Val Lys Leu Ile Glu Glu
165 170 175

Cys Thr Asp Leu Gln Lys Glu Gly Glu Phe Ser Thr Cys Phe Thr Glu
180 185 190

Leu Gln Arg Asp Phe Leu Lys Gln Arg Pro Thr Lys Leu Lys Ser Leu
195 200 205

Ile Arg Leu Val Lys His Trp Tyr Gln Asn Cys Lys Lys Lys Leu Gly
210 215 220

Lys Leu Pro Pro Gln Tyr Ala Leu Glu Leu Leu Thr Val Tyr Ala Trp
225 230 235 240

Glu Arg Gly Ser Met Lys Thr His Phe Asn Thr Ala Gln Gly Phe Arg
245 250 255

Thr Val Leu Glu Leu Val Ile Asn Tyr Gln Gln Leu Cys Ile Tyr Trp
260 265 270

Thr Lys Tyr Tyr Asp Phe Lys Asn Pro Ile Ile Glu Lys Tyr Leu Arg

275 280 285

Arg Gln Leu Thr Lys Pro Arg Pro Val Ile Leu Asp Pro Ala Asp Pro
290 295 300

Thr Gly Asn Leu Gly Gly Gly Asp Pro Lys Gly Trp Arg Gln Leu Ala
305 310 315 320

Gln Glu Ala Glu Ala Trp Leu Asn Tyr Pro Cys Phe Lys Asn Trp Asp
325 330 335

Gly Ser Pro Val Ser Ser Trp Ile Leu Leu Val Arg Pro Pro Ala Ser
340 345 350

Ser Leu Pro Phe Ile Pro Ala Pro Leu His Glu Ala
355 360

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2107 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGTAAAAGTC CACAGTTACC GTGAGAGAAA AAAAGAGGAG AAAGCAGTGC AGCCAAACTC
60

GGAAGAAAAG AGAGGAGGAA AAGGACTCGA CTTTCACATT GGAACAACCT TCTTTCCAGT
120

GCTAAGGCTC TCTGATCTGG GGAACAACAC CTGGACATGG CTCCAGAGAT CAACTTGCCG
180

GGCCCAATGA GCCTCATTGA TAACACTAAA GGGCAGCTGG TGGTGAATCC AGAAGCTCTG
240

AAGATCCTAT CTGCAATTAC GCAGCCTGTG GTGGTGGTGG CGATTGTGGG CCTCTATCGC
300

ACAGGCAAAT CCTACCTGAT GAACAAGCTG GCTGGGAAGA AAAACGGCTT CTCTCTAGGC
360

TCCACAGTGA AGTCTCACAC CAAGGGAATC TGGATGTGGT GTGTGCCTCA TCCCAAGAAG
420

CCAGAACACA CCCTAGTTCT GCTCGACACT GAGGGCCTGG GAGATATAGA GAAGGGTGAC
480

AATGAGAATG ACTCCTGGAT CTTTGCCTTG GCCATCCTCC TGAGCAGCAC CTTCGTGTAC
540

AATAGCATGG GAACCATCAA CCAGCAGGCC ATGGACCAAC TTCACTATGT GACAGAGCTG
600

ACAGATCGAA TCAAGGCAAA CTCCTCACCT GGTAACAATT CTGTAGACGA CTCAGCTGAC
660

TTTGTGAGCT TTTTTCCAGC ATTTGTGTGG ACTCTCAGAG ATTTCACCCT GGAACTGGAA
720

GTAGATGGAG AACCCATCAC TGCTGATGAC TACTTGGAGC TTTCGCTAAA GCTAAGAAAA
780

GGTACTGATA AGAAAAGTAA AAGCTTTAAT GATCCTCGGT TGTGCATCCG AAAGTTCTTC
840

CCCAAGAGGA AGTGCTTCGT CTTCGATTGG CCCGCTCCTA AGAAGTACCT TGCTCACCTA
900

GAGCAGCTAA AGGAGGAAGA GCTGAACCCT GATTTCATAG AACAAGTTGC AGAATTTTGT
960

TCCTACATCC TCAGCCATTC CAATGTCAAG ACTCTTTCAG GTGGCATTGC AGTCAATGGG
1020

CCTCGTCTAG AGAGCCTGGT GCTGACCTAC GTCAATGCCA TCAGCAGTGG GGATCTACCC
1080

TGCATGGAGA ACGCAGTCCT GGCCTTGGCC CAGATAGAGA ACTCAGCCGC AGTGGAAAAG
1140

GCTATTGCCC ACTATGAACA GCAGATGGGC CAGAAGGTGC AGCTGCCCAC GGAAACCCTC
1200

CAGGAGCTGC TGGACCTGCA CAGGGACAGT GAGAGAGAGG CCATTGAAGT CTTCATGAAG
1260

AACTCTTTCA AGGATGTGGA CCAAATGTTC CAGAGGAAAT TAGGGGCCCA GTTGGAAGCA
1320

AGGCGAGATG ACTTTTGTAA GCAGAATTCC AAAGCATCAT CAGATTGTTG CATGGCTTTA
1380

CTTCAGGATA TATTTGGCCC TTTAGAAGAA GATGTCAAGC AGGGAACATT TTCTAAACCA
1440

GGAGGTTACC GTCTCTTTAC TCAGAAGCTG CAGGAGCTGA AGAATAAGTA CTACCAGGTG
1500

CCAAGGAAGG GGATACAGGC CAAAGAGGTG CTGAAAAAAT ATTTGGAGTC CAAGGAGGAT
1560

GTGGCTGATG CACTTCTACA GACTGATCAG TCACTCTCAG AAAAGGAAAA AGCGATTGAA
1620

GTGGAACGTA TAAAGGCTGA ATCTGCAGAA GCTGCAAAGA AAATGTTGGA GGAAATACAA
1680

AAGAAGAATG AGGAGATGAT GGAACAGAAA GAGAAGAGTT ATCAGGAACA TGTGAAACAA
1740

TTGACTGAGA AGATGGAGAG GGACAGGGCC CAGTTAATGG CAGAGCAAGA GAAGACCCTC
1800

GCTCTTAAAC TTCAGGAACA GGAACGCCTT CTCAAGGAGG GATTCGAGAA TGAGAGCAAG
1860

AGACTTCAAA AAGACATATG GGATATCCAG ATGAGAAGCA AATCATTGGA GCCAATATGT
1920

AACATACTCT AAAAGTCCAA GGAGCAAAAT TTGCCTGTCC AGCTCCCTCT CCCCAAGAAA
1980

CAACATGAAT GAGCAACTTC AGAGTGTCAA ACAACTGCCA TTAAACTTAA CTCAAAATCA
2040

TGATGCATGC ATTTTTGTTG AACCATAAAG TTTGCAAAGT AAAGGTTAAG TATGAGGTCA
2100

ATGTTTT
2107

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 591 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Ala Pro Glu Ile Asn Leu Pro Gly Pro Met Ser Leu Ile Asp Asn
1               5                   10                  15

Thr Lys Gly Gln Leu Val Val Asn Pro Glu Ala Leu Lys Ile Leu Ser
            20                  25                  30

Ala Ile Thr Gln Pro Val Val Val Val Ala Ile Val Gly Leu Tyr Arg
        35                  40                  45

Thr Gly Lys Ser Tyr Leu Met Asn Lys Leu Ala Gly Lys Lys Asn Gly
    50                  55                  60

Phe Ser Leu Gly Ser Thr Val Lys Ser His Thr Lys Gly Ile Trp Met
65                  70                  75                  80

Trp Cys Val Pro His Pro Lys Lys Pro Glu His Thr Leu Val Leu Leu
                85                  90                  95

Asp Thr Glu Gly Leu Gly Asp Ile Glu Lys Gly Asp Asn Glu Asn Asp
            100                 105                 110
```

Ser Trp Ile Phe Ala Leu Ala Ile Leu Leu Ser Ser Thr Phe Val Tyr
115 120 125

Asn Ser Met Gly Thr Ile Asn Gln Gln Ala Met Asp Gln Leu His Tyr
130 135 140

Val Thr Glu Leu Thr Asp Arg Ile Lys Ala Asn Ser Ser Pro Gly Asn
145 150 155 160

Asn Ser Val Asp Asp Ser Ala Asp Phe Val Ser Phe Phe Pro Ala Phe
165 170 175

Val Trp Thr Leu Arg Asp Phe Thr Leu Glu Leu Glu Val Asp Gly Glu
180 185 190

Pro Ile Thr Ala Asp Asp Tyr Leu Glu Leu Ser Leu Lys Leu Arg Lys
195 200 205

Gly Thr Asp Lys Lys Ser Lys Ser Phe Asn Asp Pro Arg Leu Cys Ile
210 215 220

Arg Lys Phe Phe Pro Lys Arg Lys Cys Phe Val Phe Asp Trp Pro Ala
225 230 235 240

Pro Lys Lys Tyr Leu Ala His Leu Glu Gln Leu Lys Glu Glu Glu Leu
245 250 255

Asn Pro Asp Phe Ile Glu Gln Val Ala Glu Phe Cys Ser Tyr Ile Leu
260 265 270

Ser His Ser Asn Val Lys Thr Leu Ser Gly Gly Ile Ala Val Asn Gly
275 280 285

Pro Arg Leu Glu Ser Leu Val Leu Thr Tyr Val Asn Ala Ile Ser Ser
290 295 300

Gly Asp Leu Pro Cys Met Glu Asn Ala Val Leu Ala Leu Ala Gln Ile

305 310 315 320

Glu Asn Ser Ala Ala Val Glu Lys Ala Ile Ala His Tyr Glu Gln Gln
325 330 335

Met Gly Gln Lys Val Gln Leu Pro Thr Glu Thr Leu Gln Glu Leu Leu
340 345 350

Asp Leu His Arg Asp Ser Glu Arg Glu Ala Ile Glu Val Phe Met Lys
355 360 365

Asn Ser Phe Lys Asp Val Asp Gln Met Phe Gln Arg Lys Leu Gly Ala
370 375 380

Gln Leu Glu Ala Arg Arg Asp Asp Phe Cys Lys Gln Asn Ser Lys Ala
385 390 395 400

Ser Ser Asp Cys Cys Met Ala Leu Leu Gln Asp Ile Phe Gly Pro Leu
405 410 415

Glu Glu Asp Val Lys Gln Gly Thr Phe Ser Lys Pro Gly Gly Tyr Arg
420 425 430

Leu Phe Thr Gln Lys Leu Gln Glu Leu Lys Asn Lys Tyr Tyr Gln Val
435 440 445

Pro Arg Lys Gly Ile Gln Ala Lys Glu Val Leu Lys Lys Tyr Leu Glu
450 455 460

Ser Lys Glu Asp Val Ala Asp Ala Leu Leu Gln Thr Asp Gln Ser Leu
465 470 475 480

Ser Glu Lys Glu Lys Ala Ile Glu Val Glu Arg Ile Lys Ala Glu Ser
485 490 495

Ala Glu Ala Ala Lys Lys Met Leu Glu Glu Ile Gln Lys Lys Asn Glu
500 505 510

Glu Met Met Glu Gln Lys Glu Lys Ser Tyr Gln Glu His Val Lys Gln
515 520 525

Leu Thr Glu Lys Met Glu Arg Asp Arg Ala Gln Leu Met Ala Glu Gln
530 535 540

Glu Lys Thr Leu Ala Leu Lys Leu Gln Glu Gln Glu Arg Leu Leu Lys
545 550 555 560

Glu Gly Phe Glu Asn Glu Ser Lys Arg Leu Gln Lys Asp Ile Trp Asp
565 570 575

Ile Gln Met Arg Ser Lys Ser Leu Glu Pro Ile Cys Asn Ile Leu
580 585 590

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2056 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTGGAAACCT CTTCAGCATT TGCTTGGAAT CAGTAAGCTA AAAACAAAAT CAACCGGGAC
60

CCCAGCTTTT CAGAACTGCA GGGAAACAGC CATCATGAGT GAGGTCACCA AGAATTCCCT
120

GGAGAAAATC CTCCCACAGC TGAAATGCCA TTTCACCTGG AACTTATTCA AGGAAGACAG
180

TGTCTCAAGG GATCTAGAAG ATAGAGTGTG TAACCAGATT GAATTTTTAA ACACTGAGTT
240

CAAAGCTACA ATGTACAACT TGTTGGCCTA CATAAAACAC CTAGATGGTA ACAACGAGGC
300

AGCCCTGGAA TGCTTACGGC AAGCTGAAGA GTTAATCCAG CAAGAACATG CTGACCAAGC
360

AGAAATCAGA AGTCTAGTCA CTTGGGGAAA CTACGCCTGG GTCTACTATC ACTTGGGCAG
420

ACTCTCAGAT GCTCAGATTT ATGTAGATAA GGTGAAACAA ACCTGCAAGA AATTTTCAAA
480

TCCATACAGT ATTGAGTATT CTGAACTTGA CTGTGAGGAA GGGTGGACAC AACTGAAGTG
540

TGGAAGAAAT GAAAGGGCGA AGGTGTGTTT TGAGAAGGCT CTGGAAGAAA AGCCCAACAA
600

CCCAGAATTC TCCTCTGGAC TGGCAATTGC GATGTACCAT CTGGATAATC ACCCAGAGAA
660

ACAGTTCTCT ACTGATGTTT TGAAGCAGGC CATTGAGCTG AGTCCTGATA ACCAATACGT
720

CAAGGTTCTC TTGGGCCTGA AACTGCAGAA GATGAATAAA GAAGCTGAAG GAGAGCAGTT
780

TGTTGAAGAA GCCTTGGAAA AGTCTCCTTG CCAAACAGAT GTCCTCCGCA GTGCAGCCAA
840

ATTTTACAGA AGAAAAGGTG ACCTAGACAA AGCTATTGAA CTGTTTCAAC GGGTGTTGGA
900

ATCCACACCA AACAATGGCT ACCTCTATCA CCAGATTGGG TGCTGCTACA AGGCAAAAGT
960

AAGACAAATG CAGAATACAG GAGAATCTGA AGCTAGTGGA AATAAAGAGA TGATTGAAGC
1020

ACTAAAGCAA TATGCTATGG ACTATTCGAA TAAAGCTCTT GAGAAGGGAC TGAATCCTCT
1080

GAATGCATAC TCCGATCTCG CTGAGTTCCT GGAGACGGAA TGTTATCAGA CACCATTCAA
1140

TAAGGAAGTC CCTGATGCTG AAAAGCAACA ATCCCATCAG CGCTACTGCA ACCTTCAGAA
1200

ATATAATGGG AAGTCTGAAG ACACTGCTGT GCAACATGGT TTAGAGGGTT TGTCCATAAG
1260

CAAAAAATCA ACTGACAAGG AAGAGATCAA AGACCAACCA CAGAATGTAT CCGAAAATCT
1320

GCTTCCACAA AATGCACCAA ATTATTGGTA TCTTCAAGGA TTAATTCATA AGCAGAATGG
1380

AGATCTGCTG CAAGCAGCCA AATGTTATGA GAAGGAACTG GGCCGCCTGC TAAGGGATGC
1440

CCCTTCAGGC ATAGGCAGTA TTTTCCTGTC AGCATCTGAG CTTGAGGATG GTAGTGAGGA
1500

AATGGGCCAG GGCGCAGTCA GCTCCAGTCC CAGAGAGCTC CTCTCTAACT CAGAGCAACT 1560

GAACTGAGAC AGAGGAGGAA AACAGAGCAT CAGAAGCCTG CAGTGGTGGT TGTGACGGGT 1620

AGGAGGATAG GAAGACAGGG GGCCCCAACC TGGGATTGCT GAGCAGGGAA GCTTTGCATG 1680

TTGCTCTAAG GTACATTTTT AAAGAGTTGT TTTTTGGCCG GGCGCAGTGG CTCATGCCTG 1740

TAATCCCAGC ACTTTGGGAG GCCGAGGTGG GCGGATCACG AGGTCTGGAG TTTGAGACCA 1800

TCCTGGCTAA CACAGTGAAA TCCCGTCTCT ACTAAAAATA CAAAAAATTA GCCAGGCGTG 1860

GTGGCTGGCA CCTGTAGTCC CAGCTACTTG GGAGGCTGAG GCAGGAGAAT GGCGTGAACC 1920

TGGAAGGAAG AGGTTGCAGT GAGCCAAGAT TGCGCCCCTG CACTCCAGCC TGGGCAACAG 1980

AGCAAGACTC GGAATTCCTG CAGCCCGGGG GATCCACTAT TCTAGAGCGC CGCAACGGCC 2040

GTGGAGTCCA GAGATG 2056

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 490 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Ser Glu Val Thr Lys Asn Ser Leu Glu Lys Ile Leu Pro Gln Leu
1               5                   10                  15

Lys Cys His Phe Thr Trp Asn Leu Phe Lys Glu Asp Ser Val Ser Arg
            20                  25                  30

Asp Leu Glu Asp Arg Val Cys Asn Gln Ile Glu Phe Leu Asn Thr Glu
        35                  40                  45

Phe Lys Ala Thr Met Tyr Asn Leu Leu Ala Tyr Ile Lys His Leu Asp
    50                  55                  60

Gly Asn Asn Glu Ala Ala Leu Glu Cys Leu Arg Gln Ala Glu Glu Leu
65                  70                  75                  80

Ile Gln Gln Glu His Ala Asp Gln Ala Glu Ile Arg Ser Leu Val Thr
                85                  90                  95

Trp Gly Asn Tyr Ala Trp Val Tyr Tyr His Leu Gly Arg Leu Ser Asp
            100                 105                 110

Ala Gln Ile Tyr Val Asp Lys Val Lys Gln Thr Cys Lys Lys Phe Ser
        115                 120                 125

Asn Pro Tyr Ser Ile Glu Tyr Ser Glu Leu Asp Cys Glu Glu Gly Trp
    130                 135                 140
```

Thr Gln Leu Lys Cys Gly Arg Asn Glu Arg Ala Lys Val Cys Phe Glu
145 150 155 160

Lys Ala Leu Glu Glu Lys Pro Asn Asn Pro Glu Phe Ser Ser Gly Leu
165 170 175

Ala Ile Ala Met Tyr His Leu Asp Asn His Pro Glu Lys Gln Phe Ser
180 185 190

Thr Asp Val Leu Lys Gln Ala Ile Glu Leu Ser Pro Asp Asn Gln Tyr
195 200 205

Val Lys Val Leu Leu Gly Leu Lys Leu Gln Lys Met Asn Lys Glu Ala
210 215 220

Glu Gly Glu Gln Phe Val Glu Glu Ala Leu Glu Lys Ser Pro Cys Gln
225 230 235 240

Thr Asp Val Leu Arg Ser Ala Ala Lys Phe Tyr Arg Arg Lys Gly Asp
245 250 255

Leu Asp Lys Ala Ile Glu Leu Phe Gln Arg Val Leu Glu Ser Thr Pro
260 265 270

Asn Asn Gly Tyr Leu Tyr His Gln Ile Gly Cys Cys Tyr Lys Ala Lys
275 280 285

Val Arg Gln Met Gln Asn Thr Gly Glu Ser Glu Ala Ser Gly Asn Lys
290 295 300

Glu Met Ile Glu Ala Leu Lys Gln Tyr Ala Met Asp Tyr Ser Asn Lys
305 310 315 320

Ala Leu Glu Lys Gly Leu Asn Pro Leu Asn Ala Tyr Ser Asp Leu Ala
325 330 335

Glu Phe Leu Glu Thr Glu Cys Tyr Gln Thr Pro Phe Asn Lys Glu Val

340 345 350

Pro Asp Ala Glu Lys Gln Gln Ser His Gln Arg Tyr Cys Asn Leu Gln
355 360 365

Lys Tyr Asn Gly Lys Ser Glu Asp Thr Ala Val Gln His Gly Leu Glu
370 375 380

Gly Leu Ser Ile Ser Lys Lys Ser Thr Asp Lys Glu Glu Ile Lys Asp
385 390 395 400

Gln Pro Gln Asn Val Ser Glu Asn Leu Leu Pro Gln Asn Ala Pro Asn
405 410 415

Tyr Trp Tyr Leu Gln Gly Leu Ile His Lys Gln Asn Gly Asp Leu Leu
420 425 430

Gln Ala Ala Lys Cys Tyr Glu Lys Glu Leu Gly Arg Leu Leu Arg Asp
435 440 445

Ala Pro Ser Gly Ile Gly Ser Ile Phe Leu Ser Ala Ser Glu Leu Glu
450 455 460

Asp Gly Ser Glu Glu Met Gly Gln Gly Ala Val Ser Ser Ser Pro Arg
465 470 475 480

Glu Leu Leu Ser Asn Ser Glu Gln Leu Asn
485 490

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGCGTGTCTA CGCGGACGCA CCGGCTAAGC TGCTTCTGCC GCCGCCGGCC GCCTGGGACC 60

TTGCGGTGAG GCTGCGCGGG GCCGAGGCCG CCTCCGAGCG CCAGGTTTAT TCAGTCACCA 120

TGAAGCTGCT GCTGCTGCAC CCGGCCTTCC AGAGCTGCCT CCTGCTGACC CTGCTTGGCT 180

TATGGAGAAC CACCCCTGAG GCTCACGCTT CATCCCTGGG TGCACCAGCT ATCAGCGCTG 240

CCTCCTTCCT GCAGGATCTA ATACATCGGT ATGGCGAGGG TGACAGCCTC ACTCTGCAGC 300

AGCTGAAGGC CCTGCTCAAC CACCTGGATG TGGGAGTGGG CCGGGGTAAT GTCACCCAGC 360

ACGTGCAAGG ACACAGGAAC CTCTCCACGT GCTTTAGTTC TGGAGACCTC TTCACTGCCC 420

ACAATTTCAG CGAGCAGTCG CGGATTGGGA GCAGCGAGCT CCAGGAGTTC TGCCCCACCA 480

TCCTCCAGCA GCTGGATTCC CGGGCCTGCA CCTCGGAGAA CCAGGAAAAC GAGGAGAATG 540

AGCAGACGGA GGAGGGGCGG CCAAGCGCTG TTGAAGTGTG GGGATACGGT CTCCTCTGTG
600

TGACCGTCAT CTCCCTCTGC TCCCTCCTGG GGGCCAGCGT GGTGCCCTTC ATGAAGAAGA
660

CCTTTTACAA GAGGCTGCTG CTCTACTTCA TAGCTCTGGC GATTGGAACC CTCTACTCCA
720

ACGCCCTCTT CCAGCTCATC CCGGAGGCAT TTGGTTTCAA CCCTCTGGAA GATTATTATG
780

TCTCCAAGTC TGCAGTGGTG TTTGGGGGCT TTTATCTTTT CTTTTTCACA GAGAAGATCT
840

TGAAGATTCT TCTTAAGCAG AAAAATGAGC ATCATCATGG ACACAGCCAT TATGCCTCTG
900

AGTCGCTTCC CTCCAAGAAG GACCAGGAGG AGGGGGTGAT GGAGAAGCTG CAGAACGGGG
960

ACCTGGACCA CATGATTCCT CAGCACTGCA GCAGTGAGCT GGACGGCAAG GCGCCCATGG
1020

TGGACGAGAA GGTCATTGTG GGCTCGCTCT CTGTGCAGGA CCTGCAGGCT TCCCAGAGTG
1080

CTTGCTACTG GCTGAAAGGT GTCCGCTACT CTGATATCGG CACTCTGGCC TGGATGATCA
1140

CTCTGAGCGA CGGCCTCCAC AATTTCATCG ATGGCCTGGC CATCGGTGCT TCCTTCACTG
1200

TGTCAGTTTT CCAAGGCATC AGCACCTCGG TGGCCATCCT CTGTGAGGAG TTCCCACATG
1260

AGCTAGGAGA CTTTGTCATC CTGCTCAACG CTGGGATGAG CATCCAACAA GCTCTCTTCT
1320

TCAACTTCCT TTCTGCCTGC TGCTGCTACC TGGGTCTGGC CTTTGGCATC CTGGCCGGCA
1380

GCCACTTCTC TGCCAACTGG ATTTTTGCGC TAGCTGGAGG AATGTTCTTG TATATTTCTC
1440

TGGCTGATAT GTTCCCTGAG ATGAATGAGG TCTGTCAAGA GGATGAAAGG AAGGGCAGCA
1500

TCTTGATTCC ATTTATCATC CAGAACCTGG GCCTCCTGAC TGGATTCACC ATCATGGTGG
1560

TCCTCACCAT GTATTCAGGA CAGATCCAGA TTGGGTAGGG CTCTGCCAAG AGCCTGTGGG
1620

ACTGGAAGTC GGGCCCTGGG CTGCCCGATC GCCAGCCCGA GGACTTACCA TCCACAATGC
1680

ACCACGGAAG AGGCCGTTCT ATGAAAAACT GACACAGACT GTATTCCTGC ATTCAAATGT
1740

CAGCCGTTTG TAAAATGCTG TATCCTAGGA ATAAGCTGCC CTGGTAACCA GTCTCTAGCT
1800

AGTGCCTCTT GCCCTCTCCT CACCTCCTTT TCTCTCAGTG ACTCTGGAAC CTGAATGCAG
1860

CTTACAAGAC AAGCCTGACT TTTTTCTCTG ATTACCTTGG CCTCCTCTTG GAACCAGTGC
1920

TGAAAGGTTT TGAATCCTTT ACCCAACAAT GCAAAAATAG AGCCAATGGT TATAACTTGG
1980

CTAGAAATAT CAAGAGTTGA ATCCATAGTG TGGGGCCCAT GACTCTAGCT GGGCACCTTG 2040

GACCTCCAGC TGGCCAATAG AAGAGACAGG AGACAGGAAG CCTTCCCATT TTTTCAAAGT 2100

CTGTTTAATT GCCTATTACT TCTCTCAAAG AGAACCTGAA GTCAGAACAC ATGAGCAGGG 2160

TGAGAGGTGA GGCAAGGTTC ATCCTGAATG GGAGAGGAAG TCGAACCACT GCTGTGTGTC 2220

TTGTCAGGAT GCTCACTTGT TCCTACTGAG ATGCTGGATA TTGATTTTGT AACAGCACCT 2280

GGTGTTTCAC GGCTGTCCGA GTGAGCTAAC GTGGCGGTGT GGCTGCCTGG ACCTCCTCTT 2340

TCAGGTTAAC GCTGACAGAA TGGAGGCTCA GGCTGTCTGC AAGAAAACAG TTGGTTTGGC 2400

TGTGATTTTG ACCTCCTCTT CCCCACTGCC ATCTTCTAAG AGACTTTGTA GCTGCCTCCT 2460

AGAAGCACAT TCTGAGCACA TTTGAGACCT CTGTGTTAGA GGGGAGACTG CACAAACTAT 2520

CCTCCCCCAG GTTGAGACGT CTGCAGAGTG GCAAGCTGAC TTGTAGAAAT GGGGTGCCAT 2580

TTATGCTCTA CTTAGACAAG GGTAATCAGA AATGGAATCA GTGCAGGCAA AATTTAGGAT 2640

TTGCCGCTTC CATAAATCAA AGCATGACTA ATAGGGGGTC TCTGAAATGT AAGGGCACAA 2700

ACTTCACTTA GGGCATCGCA GATGTTTGCA GAATGGTTGG CCTAATGATT ATGCTACAGA
2760

TGGGTTTTAA ATGACCCGTC TAGGTTACTG CTTCCTTGCA AAAAAAGTCG AATCCTGCAT
2820

TGAATTGAAT ATGAATTTCT CTAACTCTCT CCAGAAAATG GATGGAGATA ACTTGTCTTT
2880

AAAACTGTAG GCCAGCCTTA GCCACTGTGG AGCCCTTGCC TCCGAGCTCT GGCTTCAAGG
2940

GGAGCTCTTC TCCAGGTTCA CTAGGTGAAT TGATTTATTA TTATCATATT GATAATGTGA
3000

GATTCTTTAG CCACTTTGGG GAGCCTGTCT CTCCAGAAGC CTTTCTTAGT GGTGCCCACA
3060

GTTGGAGCCC AGGGGCCATG TTTGCAAACT GATTCATGTG CATGGCTGAC AGGAGTACTG
3120

GTTCACTACC AATGCCTGAG CTTTTCTCTT ACATAGAAAA ACTGTCCACT CTCAGTAATC
3180

ACAAGCAGCA TCCGTTTTGT TTTCTCTTCT TGGGAGACAT CTGTCAAACC AGGAATATTC
3240

TTGAAAAGAA CGTGAGCAGG AAAAACTGCT GGTGATACTT TTTTTAAGTT TTGTTTTTAT
3300

CTTGCCTGTT GGCTTCAATA CATTTGAGAA TACGCTGAAG AGGGAAAATT TCAGTGATGG
3360

AGATTCTAGA TTAAATATCA GGACTGATTT CCTGGTGGGA TTATGGTCCA GTTTTACCAA
3420

AGAACCAATT CCTTGAATGT TGGAATCTAA CTTTTTATAT TGTCATTATT ATTGTTGTTT
3480

TTAAACGGTT CTTTGTCTTT TCTGTTTTAT TTTTCTCAAG CTGCTTTCAG GAGCTAGCAG
3540

AAAATAACTC AAAGTTGAAG ACTCTGGAAG ATTTTGCTTT AACCTAACTC GCATTGATGT
3600

ATTAAATTTA TAATTTTAGC ATTCCCAATA GATCCTATCA TTCCTTAAAC ATAATACCCT
3660

TTGTCTTGGA GTAGAATACT AAGTTAGAGT TAGTGGATTT CTAGTTTAGG AGAGGAGCTC
3720

AAAACTATAA TCTTTAACAA ATTGAAAAAT GAAATAGGGT GTTTTCCCTT TTTGTGCACA
3780

CCTATATTAC CTTAAGAAAT TTCCTTCCAT AGACAGCTGC CTCAAAGGGA AATCCTCTTT
3840

AAACCGTAGT TGGCGCAGAG GTCAGTCCTA GTCGGAGCTT AGGAGGGGCG GAGACGCTCA
3900

CATCGTCTGA CTTGAGTCGC CACTGATTGT GGCAACAGCT TTGCCTCATG AGTCAAAAAT
3960

TGGCAATTTC TTTTGATTTT TAGTTGTTGA ATTTGCTGTT TCAAGCATTT GTACATATTA
4020

GAAGTCTAAG GAGTAGCAAG TCAGTGGGAG GACTTTTTCA CCCCTGGCAT TAGCAGCTTC
4080

GACCTCATTT TCCAGATGCA CCAGCTCCTA TTAATAAGTT AGCAAGGAAA GTGTATGTCA
4140

CGTGCAGGAA CAGTGAGGCA GGGACAGGGG TTCTGCTCCT TCTCACTTCA CCACCGGCAC
4200

ACAGCTTGCC CCTGTCTTTG CCCCCAAAGG TATTTTGTGT CTAGTGTCAA ATTGGAGCTA
4260

TTCTTCACTG GTCCTTAACC TTGGGTTTTA AAAAGAAGGC TTCTCTGTTT GGGTAGCGTA
4320

AGAGCTGAGT ATAGTAAGTC CTCTTCCAAA GAGATGGCAA TATGCTGGGC ATCTACTTTA
4380

AAACAAAGTT GTCTGATTTT TGCAAGAGAG GTTAGGATTT TATTGTTCTT ATTTCCCTTT
4440

ACAGTTCTGC AGTTCCATCA CAGTATTTTT TTAAATAACT CAGGTGTATG AGCAGAAATT
4500

AGAAAAGAAA ATTAACTTAT GTGGACTGTA AATGTTTTAT TTGTAAGATT CTATAAATAA
4560

AGCTATATTC TGT
4573

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 531 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Arg Val Tyr Ala Asp Ala Pro Ala Lys Leu Leu Leu Pro Pro Pro Ala
1   5   10   15

Ala Trp Asp Leu Ala Val Arg Leu Arg Gly Ala Glu Ala Ala Ser Glu
            20   25   30

Arg Gln Val Tyr Ser Val Thr Met Lys Leu Leu Leu Leu His Pro Ala
        35   40   45

Phe Gln Ser Cys Leu Leu Leu Thr Leu Leu Gly Leu Trp Arg Thr Thr
    50   55   60

Pro Glu Ala His Ala Ser Ser Leu Gly Ala Pro Ala Ile Ser Ala Ala
65   70   75   80

Ser Phe Leu Gln Asp Leu Ile His Arg Tyr Gly Glu Gly Asp Ser Leu
                85   90   95

Thr Leu Gln Gln Leu Lys Ala Leu Leu Asn His Leu Asp Val Gly Val
            100   105   110

Gly Arg Gly Asn Val Thr Gln His Val Gln Gly His Arg Asn Leu Ser
        115   120   125

Thr Cys Phe Ser Ser Gly Asp Leu Phe Thr Ala His Asn Phe Ser Glu
    130   135   140

Gln Ser Arg Ile Gly Ser Ser Glu Leu Gln Glu Phe Cys Pro Thr Ile
145   150   155   160

Leu Gln Gln Leu Asp Ser Arg Ala Cys Thr Ser Glu Asn Gln Glu Asn

165 170 175

Glu Glu Asn Glu Gln Thr Glu Glu Gly Arg Pro Ser Ala Val Glu Val
180 185 190

Trp Gly Tyr Gly Leu Leu Cys Val Thr Val Ile Ser Leu Cys Ser Leu
195 200 205

Leu Gly Ala Ser Val Val Pro Phe Met Lys Lys Thr Phe Tyr Lys Arg
210 215 220

Leu Leu Leu Tyr Phe Ile Ala Leu Ala Ile Gly Thr Leu Tyr Ser Asn
225 230 235 240

Ala Leu Phe Gln Leu Ile Pro Glu Ala Phe Gly Phe Asn Pro Leu Glu
245 250 255

Asp Tyr Tyr Val Ser Lys Ser Ala Val Val Phe Gly Gly Phe Tyr Leu
260 265 270

Phe Phe Phe Thr Glu Lys Ile Leu Lys Ile Leu Leu Lys Gln Lys Asn
275 280 285

Glu His His His Gly His Ser His Tyr Ala Ser Glu Ser Leu Pro Ser
290 295 300

Lys Lys Asp Gln Glu Glu Gly Val Met Glu Lys Leu Gln Asn Gly Asp
305 310 315 320

Leu Asp His Met Ile Pro Gln His Cys Ser Ser Glu Leu Asp Gly Lys
325 330 335

Ala Pro Met Val Asp Glu Lys Val Ile Val Gly Ser Leu Ser Val Gln
340 345 350

Asp Leu Gln Ala Ser Gln Ser Ala Cys Tyr Trp Leu Lys Gly Val Arg
355 360 365

Tyr Ser Asp Ile Gly Thr Leu Ala Trp Met Ile Thr Leu Ser Asp Gly
370 375 380

Leu His Asn Phe Ile Asp Gly Leu Ala Ile Gly Ala Ser Phe Thr Val
385 390 395 400

Ser Val Phe Gln Gly Ile Ser Thr Ser Val Ala Ile Leu Cys Glu Glu
405 410 415

Phe Pro His Glu Leu Gly Asp Phe Val Ile Leu Leu Asn Ala Gly Met
420 425 430

Ser Ile Gln Gln Ala Leu Phe Phe Asn Phe Leu Ser Ala Cys Cys Cys
435 440 445

Tyr Leu Gly Leu Ala Phe Gly Ile Leu Ala Gly Ser His Phe Ser Ala
450 455 460

Asn Trp Ile Phe Ala Leu Ala Gly Gly Met Phe Leu Tyr Ile Ser Leu
465 470 475 480

Ala Asp Met Phe Pro Glu Met Asn Glu Val Cys Gln Glu Asp Glu Arg
485 490 495

Lys Gly Ser Ile Leu Ile Pro Phe Ile Ile Gln Asn Leu Gly Leu Leu
500 505 510

Thr Gly Phe Thr Ile Met Val Val Leu Thr Met Tyr Ser Gly Gln Ile
515 520 525

Gln Ile Gly
530

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3200 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CAGGAAGGGC CATGAAGATT AATAAAGATT TGGACTCAGG GCAAATATTT ACTTAGTAGC 60

AATAACTCAA AGAATTACTG TTGAATAAAT AAGCCAATTA AGCAGCCAAT CACGTACTAT 120

GCGGATGCAC ACAAATGAAA CCCTCACTTC AACCTGAAGA CATTCGCACA TGAGTTACGT 180

AGAGGGACCT GCAGGAAGCG GTAGAGAAAA CATAAGGCTT ATGCGTTTAA TTTCCACACC 240

AATTTCAGGA TCTTTGTCAC TGACAGCAGC ACTAAGACTT GTTAACTTTA TATAGTTAAG 300

AAGAACAAGG CTGAGCGCGA TGACTCACGC CTGTAAGCCT AGAACTTTGG GAGGCCAAAG 360

CAGGCAGACT GCTTGAGCCC AGGAGTTCCA GACCAGCCTG GGCAACATGG CAACACCCCA 420

TCTCTACAAA AAAATACAAG AATCAGCTGG GCGTGGTGAT GTGTTCCTGT AATCTCAGCT 480

ACTCGGGAGG CAGAGGCAGG AGGATTGCTT GAACCCGGGA GGCAGAGGTT GTAGTTAGCC 540

GAGATCTCGC CACTGCACTC CAGTCTGGAC GACAGAGTGA GACTCAGTCT CAAATAAATA 600

AATAAATACA TAAATATAAG GAAAAAAATA AAGCTGCTTT CTCCTCTTCC TCCTCTTTGG 660

TCTCATCTGG CTCTGCTCCA GGCATCTGCC ACAATGTGGG TGCTTACACC TGCTGCTTTT 720

GCTGGGAAGT TCTTGAGTGT GTTCAGGCAA CCTCTGAGCT CTCTGTGGAG GAGCCTGGTC 780

CCGCTGTTCT GCTGGCTGAG GGCAACCTTC TGGCTGCTAG CTACCAAGAG GAGAAAGCAG 840

CAGCTGGTCC TGAGAGGGCC AGATGAGACC AAAGAGGAGG AAGAGGACCC TCCTCTGCCC 900

ACCACCCCAA CCAGCGTCAA CTATCACTTC ACTCGCCAGT GCAACTACAA ATGCGGCTTC 960

TGTTTCCACA CAGCCAAAAC ATCCTTTGTG CTGCCCCTTG AGGAAGCAAA GAGAGGATTG 1020

CTTTTGCTTA AGGAAGCTGG TATGGAGAAG ATCAACTTTT CAGGTGGAGA GCCATTTCTT 1080

CAAGACCGGG GAGAATACCT GGGCAAGTTG GTGAGGTTCT GCAAAGTAGA GTTGCGGCTG 1140

CCCAGCGTGA GCATCGTGAG CAATGGAAGC CTGATCCGGG AGAGGTGGTT CCAGAATTAT
1200

GGTGAGTATT TGGACATTCT CGCTATCTCC TGTGACAGCT TTGACGAGGA AGTCAATGTC
1260

CTTATTGGCC GTGGCCAAGG AAAGAAGAAC CATGTGGAAA ACCTTCAAAA GCTGAGGAGG
1320

TGGTGTAGGG ATTATAGAAT CCCTTTCAAG ATAAATTCTG TCATTAATCG TTTCAACGTG
1380

GAAGAGGACA TGACGGAACA GATCAAAGCA CTAAACCCTG TCCGCTGGAA AGTGTTCCAG
1440

TGCCTCTTAA TTGAAGGTGA GAATTGTGGA GAAGATGCTC TAAGAGAAGC AGAAAGATTT
1500

GTTATTGGTG ATGAAGAATT TGAAAGATTC TTGGAGCGCC ACAAAGAAGT GTCCTGCTTG
1560

GTGCCTGAAT CTAACCAGAA GATGAAAGAC TCCTACCTTA TTCTGGATGA ATATATGCGC
1620

TTTCTGAACT GTAGAAAGGG ACGGAAGGAC CCTTCCAAGT CCATCCTGGA TGTTGGTGTA
1680

GAAGAAGCTA TAAAATTCAG TGGATTTGAT GAAAAGATGT TTCTGAAGCG AGGAGGAAAA
1740

TACATATGGA GTAAGGCTGA TCTGAAGCTG GATTGGTAGA GCGGAAAGTG GAACGAGACT
1800

TCAACACACC AGTGGGAAAA CTCCTAGAGT AACTGCCATT GTCTGCAATA CTATCCCGTT
1860

GGTATTTCCC AGTGGCTGAA AACCTGATTT TCTGCTGCAC GTGGCATCTG ATTACCTGTG 1920

GTCACTGAAC ACACGAATAA CTTGGATAGC AAATCCTGAG ACAATGGAAA ACCATTAACT 1980

TTACTTCATT GGCTTATAAC CTTGTTGTTA TTGAAACAGC ACTTCTGTTT TTGAGTTTGT 2040

TTTAGCTAAA AAGAAGGAAT ACACACAGGA ATAATGACCC CAAAAATGCT TAGATAAGGC 2100

CCCTATACAC AGGACCTGAC ATTTAGCTCA ATGATGCGTT TGTAAGAAAT AAGCTCTAGT 2160

GATATCTGTG GGGGCAATAT TTAATTTGGA TTTGATTTTT TAAAACAATG TTTACTGCGA 2220

TTTCTATATT TCCATTTTGA AACTATTTCT TGTTCCAGGT TTGTTCATTT GACAGAGTCA 2280

GTATTTTTTG CCAAATATCC AGATAACCAG TTTTCACATC TGAGACATTA CAAAGTATCT 2340

GCCTCAATTA TTTCTGCTGG TTATAATGCT TTTTTTTTTT TTTGCTTTTA TGCCATTGCA 2400

GTCTTGTACT TTTTACTGTG ATGTACAGAA ATAGTCAACA GATGTTTCCA AGAACATATG 2460

ATATGATAAT CCTACCAATT TTCAAGAAGT CTCTAGAAAG AGATAACACA TGGAAAGACG 2520

GCGTGGTGCA GCCCAGCCCA CGGTGCCTGT TCCATGAATG CTGGCTACCT ATGTGTGTGG 2580

TACCTGTTGT GTCCCTTTCT CTTCAAAGAT CCCTGAGCAA AACAAAGATA CGCTTTCCAT
2640

TTGATGATGG AGTTGACATG GAGGCAGTGC TTGCATTGCT TTGTTCGCCT ATCATCTGGC
2700

CACATGAGGC TGTCAAGCAA AAGAATAGGA GTGTAGTTGA GTAGCTGGTT GGCCCTACAT
2760

TTCTGAGAAG TGACGTTACA CTGGGTTGGC ATAAGATATC CTAAAATCAC GCTGGAACCT
2820

TGGGCAAGGA AGAATGTGAG CAAGAGTAGA GAGAGTGCCT GGATTTCATG TCAGTGAAGC
2880

CATGTCACCA TATCATATTT TTGAATGAAC TCTGAGTCAG TTGAAATAGG GTACCATCTA
2940

GGTCAGTTTA AGAAGAGTCA GCTCAGAGAA AGCAAGCATA AGGGAAAATG TCACGTAAAC
3000

TAGATCAGGG AACAAAATCC TCTCCTTGTG GAAATATCCC ATGCAGTTTG TTGATACAAC
3060

TTAGTATCTT ATTGCCTAAA AAAAAATTTC TTATCATTGT TTCAAAAAAG CAAAATCATG
3120

GAAAATTTTT GTTGTCCAGG CAAATAAAAG GTCATTTTAA TTTAAAAAAA AAAAAAAAAA
3180

AAAAAAAAAA AAAAAGGCCA
3200

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 324 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AGGAAAAAAA ATATTCCTAC TTAAATTTTA AGTCTATAAT TCAATTTAAA TATGTGTGTG
60

TCTCATCCAG GATAGGATAG GTTGTCTTCT ATTTTCCATT TTACCTATTT ACTTTTTTTG
120

TAAGAAAAGA GAAGAATGAA TTCTAAAGAT GTTCCCCATG GGTTTTGATT GTGTCTAAGC
180

TATGATGACC TTCATATAAT CAGCATAAAC ATAAAACAAA TTTTTTACTT AACATGAGTG
240

CACTTTACTA ATCCTCATGG CACAGTGGCT CACGCCTGTA ATCCCAGCAC TTGGGGAGGA
300

CAATGTGGGG TGGATCACGA GGTC
324

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 456 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CATCTCTGGA CTCANGGCCG TTNCGGNCGC TCTANAATAG TGCATCCCCC GGGCTGCAGG
60

AATTCGGCAC GTTATAGTTC ATTACAGTTA CATAGTCCGA AGGTCTTACA ACCTAATCAC
120

TGGTAGCAAT AAATGCTTCA GGCCCACATG ATGCTGATTA GTTCTCAGTT TTCATTCAGT
180

TCACAATATA ACCACCATTC CTGCCCTCCC TGCCAAGGGT CATAAATGGT GACTGCCTAA
240

CAACAAAATT TGCAGTCTCA TCTCATTTTC ATCCAGACTT CTGGAACTCA AAGATTAACT
300

TTTGACTAAC CCTGGAATAT CTCTTATCTC ACTTATAGCT TCAGGCATGT ATTTATATGT
360

ATTCTTGATA GCAATACCAT AATCAATGTG TATTCCTGAT AGTAATGCTA CAATAAATCC
420

AAACATTTCA ACTCTGTTAA AAAAAAAAAA AAAAAA

456

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 397 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TTAACCCTCA CTAATGCTGG GTGACCAAAG TCTAAATAGG GCTCAGTATC CCCCATCGCT

60

TATCTCTGCC TCCTTCCTCC TCCTTCCCAG TCTATCATCA ACCTTGAGTA TTCTACACAA

120

TGTGAATTCG AGTGCCTGAT TAATTGAGGT GGCAACATAG TTTGAGACGA GGGCAGAGAA

180

CAGGAAGATA CATAGCTAGA AGCGACGGGT ACAAAAAGCA ATGTGTACAA GAAGACTTTC

240

AGCAAGTATA CAGAGAGTTC ACCTCTACTC TGCCCTCCTC ATAGTCATAA TGTAGCAAGT

300

AAAGAATGAG AATGGATTCT GTACAATACA CTAGAAACCA ACATAATGTA TTTCTTTAAA

360

ACCTGTGTGA AAAAAAAAGA TATCACTCAG CATAATG

397

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 272 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCGATTATT AACCCTCACT AAAGCACCGT CCAAGTTTAT TTGTGGCATT TTATGACTAC

60

CAAGCATACC CAGAGTACCT TATTACGTTT AGAAAATAAC ACTTTGGTAT CCTTCCCACA

120

AAATTATTCT CCATTTGTAC ATATCTAGTT GTAAAACAAG TTTTAGCTTT TTTTTTTAAT

180

TCCTCTTAAC AGATTTTTCT AATATCCAAG GATCATTCTT TGTCGCTGCA GTCAGATCTT

240

TAGTGAGGGT TAATAATCCA TATGACTAGT AG
272

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2651 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGAATTCTGT GGCCATACTG CGAGGAGATC GGTTCCGGGT CGGAGGCTAC AGGAAGACTC
60

CCACTCCCTG AAATCTGGAG TGAAGAACGC CGCCATCCAG CCACCATTCC AAGGAGGTGC
120

AGGAGAACAG CTCTGTGATA CCATTTAACT TGTTGACATT ACTTTTATTT GAAGGAACGT
180

ATATTAGAGC TTACTTTGCA AAGAAGGAAG ATGGTTGTTT CCGAAGTGGA CATCGCAAAA
240

GCTGATCCAG CTGCTGCATC CCACCCTCTA TTACTGAATG GAGATGCTAC TGTGGCCCAG
300

AAAAATCCAG GCTCGGTGGC CGAGAACAAC CTGTGCAGCC AGTATGAGGA GAAGGTGCGC 360

CCCTGCATCG ACCTCATTGA CTCCCTGCGG GCTCTAGGTG TGGAGCAGGA CCTGGCCCTG 420

CCAGCCATCG CCGTCATCGG GGACCAGAGC TCGGGCAAGA GCTCCGTGTT GGAGGCACTG 480

TCAGGAGTTG CCCTTCCCAG AGGCAGCGGG ATCGTGACCA GATGCCCGCT GGTGCTGAAA 540

CTGAAGAAAC TTGTGAACGA AGATAAGTGG AGAGGCAAGG TCAGTTACCA GGACTACGAG 600

ATTGAGATTT CGGATGCTTC AGAGGTAGAA AAGGAAATTA ATAAAGCCCA GAATGCCATC 660

GCCGGGGAAG GAATGGGAAT CAGTCATGAG CTAATCACCC GTGAGATCAG CTCCCGAGAT 720

GTCCCGGATC TGACTCTAAT AGACCTTCCT GGCATAACCA GAGTGGCTGT GGGCAATCAG 780

CCTGCTGACA TTGGGTATAA GATCAAGACA CTCATCAAGA AGTACATCCA GAGGCAGGAG 840

ACAATCAGCC TGGTGGTGGT CCCCAGTAAT GTGGACATTG CCACCACAGA GGCTCTCAGC 900

ATGGCCCAGG AGGTGGACCC CGAGGGAGAC AGGACCATCG GAATCTTGAC GAAGCCTGAT 960

CTGGTGGACA AAGGAACTGA AGACAAGGTT GTGGACGTGG TGCGGAACCT CGTGTTCCAC 1020

CTGAAGAAGG GTTACATGAT TGTCAAGTGC CGGGGCCAGC AGGAGATCCA GGACCAGCTG
1080

AGCCTGTCCG AAGCCCTGCA GAGAGAGAAG ATCTTCTTTG AGAACCACCC ATATTTCAGG
1140

GATCTGCTGG AGGAAGGAAA GGCCACGGTT CCCTGCCTGG CAGAAAAACT TACCAGCGAG
1200

CTCATCACAC ATATCTGTAA ATCTCTGCCC CTGTTAGAAA ATCAAATCAA GGAGACTCAC
1260

CAGAGAATAA CAGAGGAGCT ACAAAAGTAT GGTGTCGACA TACCGGAAGA CGAAAATGAA
1320

AAAATGTTCT TCCTGATAGA TAAAATTAAT GCCTTTAATC AGGACATCAC TGCTCTCATG
1380

CAAGGAGAGG AAACTGTAGG GGAGGAAGAC ATTCGGCTGT TTACCAGACT CCGACACGAG
1440

TTCCACAAAT GGAGTACAAT AATTGAAAAC AATTTTCAAG AAGGCCATAA AATTTTGAGT
1500

AGAAAAATCC AGAAATTTGA AAATCAGTAT CGTGGTAGAG AGCTGCCAGG CTTTGTGAAT
1560

TACAGGACAT TTGAGACAAT CGTGAAACAG CAAATCAAGG CACTGGAAGA GCCGGCTGTG
1620

GATATGCTAC ACACCGTGAC GGATATGGTC CGGCTTGCTT TCACAGATGT TTCGATAAAA
1680

AATTTTGAAG AGTTTTTTAA CCTCCACAGA ACCGCCAAGT CCAAAATTGA AGACATTAGA
1740

GCAGAACAAG AGAGAGAAGG TGAGAAGCTG ATCCGCCTCC ACTTCCAGAT GGAACAGATT
1800

GTCTACTGCC AGGACCAGGT ATACAGGGGT GCATTGCAGA AGGTCAGAGA GAAGGAGCTG
1860

GAAGAAGAAA AGAAGAAGAA ATCCTGGGAT TTTGGGGCTT TCCAATCCAG CTCGGCAACA
1920

GACTCTTCCA TGGAGGAGAT CTTTCAGCAC CTGATGGCCT ATCACCAGGA GGCCAGCAAG
1980

CGCATCTCCA GCCACATCCC TTTGATCATC CAGTTCTTCA TGCTCCAGAC GTACGGCCAG
2040

CAGCTTCAGA AGGCCATGCT GCAGCTCCTG CAGGACAAGG ACACCTACAG CTGGCTCCTG
2100

AAGGAGCGGA GCGACACCAG CGACAAGCGG AAGTTCCTGA AGGAGCGGCT TGCACGGCTG
2160

ACGCAGGCTC GGCGCCGGCT TGCCCAGTTC CCCGGTTAAC CACACTCTGT CCAGCCCCGT
2220

AGACGTGCAC GCACACTGTC TGCCCCCGTT CCCGGGTAGC CACTGGACTG ACGACTTGAG
2280

TGCTCAGTAG TCAGACTGGA TAGTCCGTTC CTGCTTATCC GTTAGCCGTG GTGATTTAGC
2340

AGGAAGCTGT GAGAGCAGTT TGGTTTCTAG CATGAAGACA GAGCCCCACC CTCAGATGCA
2400

CATGAGCTGG CGGGATTGAA GGATGCTGTC TTCGTACTGG GAAAGGGATT TTCAGCCCTC
2460

AGAATCGCTC CACCTTGCAG CTCTCCCCTT CTCTGTATTC CTAGAAACTG ACACATGCTG
2520

AACATCACAG CTTATTTCCT CATTTTTATA ATGTCCCTTC ACAAACCCAG TGTTTTAGGA
2580

GCATGAGTGC CGTGTGTGTG CGTCCTGTCG GAGCCCTGTC TCTCTCTCTG TAATAAACTC
2640

ATTTCTAGCA G
2651

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 662 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met Val Val Ser Glu Val Asp Ile Ala Lys Ala Asp Pro Ala Ala Ala
1 5 10 15

Ser His Pro Leu Leu Leu Asn Gly Asp Ala Thr Val Ala Gln Lys Asn
20 25 30

Pro Gly Ser Val Ala Glu Asn Asn Leu Cys Ser Gln Tyr Glu Glu Lys
35 40 45

Val Arg Pro Cys Ile Asp Leu Ile Asp Ser Leu Arg Ala Leu Gly Val
50 55 60

Glu Gln Asp Leu Ala Leu Pro Ala Ile Ala Val Ile Gly Asp Gln Ser
65 70 75 80

Ser Gly Lys Ser Ser Val Leu Glu Ala Leu Ser Gly Val Ala Leu Pro
85 90 95

Arg Gly Ser Gly Ile Val Thr Arg Cys Pro Leu Val Leu Lys Leu Lys
100 105 110

Lys Leu Val Asn Glu Asp Lys Trp Arg Gly Lys Val Ser Tyr Gln Asp
115 120 125

Tyr Glu Ile Glu Ile Ser Asp Ala Ser Glu Val Glu Lys Glu Ile Asn
130 135 140

Lys Ala Gln Asn Ala Ile Ala Gly Glu Gly Met Gly Ile Ser His Glu
145 150 155 160

Leu Ile Thr Arg Glu Ile Ser Ser Arg Asp Val Pro Asp Leu Thr Leu
165 170 175

Ile Asp Leu Pro Gly Ile Thr Arg Val Ala Val Gly Asn Gln Pro Ala
180 185 190

Asp Ile Gly Tyr Lys Ile Lys Thr Leu Ile Lys Lys Tyr Ile Gln Arg
195 200 205

Gln Glu Thr Ile Ser Leu Val Val Val Pro Ser Asn Val Asp Ile Ala
210 215 220

Thr Thr Glu Ala Leu Ser Met Ala Gln Glu Val Asp Pro Glu Gly Asp

225 230 235 240

Arg Thr Ile Gly Ile Leu Thr Lys Pro Asp Leu Val Asp Lys Gly Thr
245 250 255

Glu Asp Lys Val Val Asp Val Val Arg Asn Leu Val Phe His Leu Lys
260 265 270

Lys Gly Tyr Met Ile Val Lys Cys Arg Gly Gln Gln Glu Ile Gln Asp
275 280 285

Gln Leu Ser Leu Ser Glu Ala Leu Gln Arg Glu Lys Ile Phe Phe Glu
290 295 300

Asn His Pro Tyr Phe Arg Asp Leu Leu Glu Glu Gly Lys Ala Thr Val
305 310 315 320

Pro Cys Leu Ala Glu Lys Leu Thr Ser Glu Leu Ile Thr His Ile Cys
325 330 335

Lys Ser Leu Pro Leu Leu Glu Asn Gln Ile Lys Glu Thr His Gln Arg
340 345 350

Ile Thr Glu Glu Leu Gln Lys Tyr Gly Val Asp Ile Pro Glu Asp Glu
355 360 365

Asn Glu Lys Met Phe Phe Leu Ile Asp Lys Ile Asn Ala Phe Asn Gln
370 375 380

Asp Ile Thr Ala Leu Met Gln Gly Glu Glu Thr Val Gly Glu Glu Asp
385 390 395 400

Ile Arg Leu Phe Thr Arg Leu Arg His Glu Phe His Lys Trp Ser Thr
405 410 415

Ile Ile Glu Asn Asn Phe Gln Glu Gly His Lys Ile Leu Ser Arg Lys
420 425 430

Ile Gln Lys Phe Glu Asn Gln Tyr Arg Gly Arg Glu Leu Pro Gly Phe
435 440 445

Val Asn Tyr Arg Thr Phe Glu Thr Ile Val Lys Gln Gln Ile Lys Ala
450 455 460

Leu Glu Glu Pro Ala Val Asp Met Leu His Thr Val Thr Asp Met Val
465 470 475 480

Arg Leu Ala Phe Thr Asp Val Ser Ile Lys Asn Phe Glu Glu Phe Phe
485 490 495

Asn Leu His Arg Thr Ala Lys Ser Lys Ile Glu Asp Ile Arg Ala Glu
500 505 510

Gln Glu Arg Glu Gly Glu Lys Leu Ile Arg Leu His Phe Gln Met Glu
515 520 525

Gln Ile Val Tyr Cys Gln Asp Gln Val Tyr Arg Gly Ala Leu Gln Lys
530 535 540

Val Arg Glu Lys Glu Leu Glu Glu Glu Lys Lys Lys Lys Ser Trp Asp
545 550 555 560

Phe Gly Ala Phe Gln Ser Ser Ser Ala Thr Asp Ser Ser Met Glu Glu
565 570 575

Ile Phe Gln His Leu Met Ala Tyr His Gln Glu Ala Ser Lys Arg Ile
580 585 590

Ser Ser His Ile Pro Leu Ile Ile Gln Phe Phe Met Leu Gln Thr Tyr
595 600 605

Gly Gln Gln Leu Gln Lys Ala Met Leu Gln Leu Leu Gln Asp Lys Asp
610 615 620

Thr Tyr Ser Trp Leu Leu Lys Glu Arg Ser Asp Thr Ser Asp Lys Arg

625 630 635 640

Lys Phe Leu Lys Glu Arg Leu Ala Arg Leu Thr Gln Ala Arg Arg Arg
645 650 655

Leu Ala Gln Phe Pro Gly
660

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 556 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATGGGCTGGG ACCTGACGGT GAAGATGCTG GCGGGCAACG AATTCCAGGT GTCCCTGAGC 60

AGCTCCATGT CGGTGTCAGA GCTGAAGGCG CAGATCACCC AGAACATTGG CGTGCACGCC 120

TTCCAGCAGC GTCTGGCTGT CCACCCGAGC GGTGTGGCGC TGCAGGACAG GGTCCCCCTT 180

GCCAGCCAGG GCCTGGGCCC TGGCAGCACG GTCCTGCTGG TGGTGGACAA ATGCGACGAA
240

CCTCTGAGCA TCCTGGTGAG GAATAACAAG GGCCGCAGCA GCACCTACGA GGTGCGGCTG
300

ACGCAGACCG TGGCCCACCT GAAGCAGCAA GTGAGCGGGC TGGAGGGTGT GCAGGACGAC
360

CTGTTCTGGC TGACCTTCGA GGGGAAGCCC CTGGAGGACC AGCTCCCGCT GGGGGAGTAC
420

GGCCTCAAGC CCCTGAGCAC CGTGTTCATG AATCTGCGCC TGCGGGGAGG CGGCACAGAG
480

CCTGGCGGGC GGAGCTAAGG GCCTCCACCA GCATCCGAGC AGGATCAAGG GCCGGAATAA
540

AGGCTGTTGT AAGAGA
556

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 165 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Gly Trp Asp Leu Thr Val Lys Met Leu Ala Gly Asn Glu Phe Gln
1 5 10 15

Val Ser Leu Ser Ser Ser Met Ser Val Ser Glu Leu Lys Ala Gln Ile
20 25 30

Thr Gln Asn Ile Gly Val His Ala Phe Gln Gln Arg Leu Ala Val His
35 40 45

Pro Ser Gly Val Ala Leu Gln Asp Arg Val Pro Leu Ala Ser Gln Gly
50 55 60

Leu Gly Pro Gly Ser Thr Val Leu Leu Val Val Asp Lys Cys Asp Glu
65 70 75 80

Pro Leu Ser Ile Leu Val Arg Asn Asn Lys Gly Arg Ser Ser Thr Tyr
85 90 95

Glu Val Arg Leu Thr Gln Thr Val Ala His Leu Lys Gln Gln Val Ser
100 105 110

Gly Leu Glu Gly Val Gln Asp Asp Leu Phe Trp Leu Thr Phe Glu Gly
115 120 125

Lys Pro Leu Glu Asp Gln Leu Pro Leu Gly Glu Tyr Gly Leu Lys Pro
130 135 140

Leu Ser Thr Val Phe Met Asn Leu Arg Leu Arg Gly Gly Gly Thr Glu
145 150 155 160

Pro Gly Gly Arg Ser
165

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1360 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGCACGAGTT CAGTTTCAGT AGCTCTGCGT GTAGAAAAGA AACGCCATGG CTGACAAGAT 60

CCTGAGGGCA AAGAGGAAGC AATTTATCAA CTCAGTGAGT ATAGGGACAA TAAATGGATT 120

GTTGGATGAA CTTTTAGAGA AGAGAGTGCT GAATCAGGAA GAAATGGATA AAATAAAACT 180

TGCAAACATT ACTGCTATGG ACAAGGCACG GGACCTATGT GATCATGTCT CTAAAAAAGG 240

GCCCCAGGCA AGCCAAATCT TTATCACTTA CATTTGTAAT GAAGACTGCT ACCTGGCAGG 300

AATTCTGGAG CTTCAATCAG CTCCATCAGC TGAAACATTT GTTGCTACAG AAGATTCTAA 360

AGGAGGACAT CCTTCATCCT CAGAAACAAA GGAAGAACAG AACAAAGAAG ATGGCACATT 420

TCCAGGACTG ACTGGGACCC TCAAGTTTTG CCCTTTAGAA AAAGCCCAGA AGTTATGGAA
480

AGAAAATCCT TCAGAGATTT ATCCAATAAT GAATACAACC ACTCGTACAC GTCTTGCCCT
540

CATTATCTGC AACACAGAGT TTCAACATCT TTCTCCGAGG GTTGGAGCTC AAGTTGACCT
600

CAGAGAAATG AAGTTGCTGC TGGAGGATCT GGGGTATACC GTGAAAGTGA AAGAAAATCT
660

CACAGCTCTG GAGATGGTGA AAGAGGTGAA AGAATTTGCT GCCTGCCCAG AGCACAAGAC
720

TTCTGACAGT ACTTTCCTTG TATTCATGTC TCATGGTATC CAGGAGGGAA TATGTGGGAC
780

CACATACTCT AATGAAGTTT CAGATATTTT AAAGGTTGAC ACAATCTTTC AGATGATGAA
840

CACTTTGAAG TGCCCAAGCT TGAAAGACAA GCCCAAGGTG ATCATTATTC AGGCATGCCG
900

TGGAGAGAAA CAAGGAGTGG TGTTGTTAAA AGATTCAGTA AGAGACTCTG AAGAGGATTT
960

CTTAACGGAT GCAATTTTTG AAGATGATGG CATTAAGAAG GCCCATATAG AGAAAGATTT
1020

TATTGCTTTC TGCTCTTCAA CACCAGATAA TGTGTCTTGG AGACATCCTG TCAGGGGCTC
1080

ACTTTTCATT GAGTCACTCA TCAAACACAT GAAAGAATAT GCCTGGTCTT GTGACTTGGA
1140

GGACATTTTC AGAAAGGTTC GATTTTCATT TGAACAACCA GAATTTAGGC TACAGATGCC
1200

CACTGCTGAT AGGGTGACCC TGACAAAACG TTTCTACCTC TTCCCGGGAC ATTAAACGAA
1260

GAATCCAGTT CATTCTTATG TACCTATGCT GAGAATCGTG CCAATAAGAA GCCAATACTT
1320

CCTTAGATGA TGCAATAAAT ATTAAAATAA AACAAAACAG
1360

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 402 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Ala Asp Lys Ile Leu Arg Ala Lys Arg Lys Gln Phe Ile Asn Ser
1 5 10 15

Val Ser Ile Gly Thr Ile Asn Gly Leu Leu Asp Glu Leu Leu Glu Lys
20 25 30

Arg Val Leu Asn Gln Glu Glu Met Asp Lys Ile Lys Leu Ala Asn Ile
35 40 45

Thr Ala Met Asp Lys Ala Arg Asp Leu Cys Asp His Val Ser Lys Lys
50 55 60

Gly Pro Gln Ala Ser Gln Ile Phe Ile Thr Tyr Ile Cys Asn Glu Asp
65 70 75 80

Cys Tyr Leu Ala Gly Ile Leu Glu Leu Gln Ser Ala Pro Ser Ala Glu
85 90 95

Thr Phe Val Ala Thr Glu Asp Ser Lys Gly Gly His Pro Ser Ser Ser
100 105 110

Glu Thr Lys Glu Glu Gln Asn Lys Glu Asp Gly Thr Phe Pro Gly Leu
115 120 125

Thr Gly Thr Leu Lys Phe Cys Pro Leu Glu Lys Ala Gln Lys Leu Trp
130 135 140

Lys Glu Asn Pro Ser Glu Ile Tyr Pro Ile Met Asn Thr Thr Thr Arg
145 150 155 160

Thr Arg Leu Ala Leu Ile Ile Cys Asn Thr Glu Phe Gln His Leu Ser
165 170 175

Pro Arg Val Gly Ala Gln Val Asp Leu Arg Glu Met Lys Leu Leu Leu
180 185 190

Glu Asp Leu Gly Tyr Thr Val Lys Val Lys Glu Asn Leu Thr Ala Leu
195 200 205

Glu Met Val Lys Glu Val Lys Glu Phe Ala Ala Cys Pro Glu His Lys
210 215 220

Thr Ser Asp Ser Thr Phe Leu Val Phe Met Ser His Gly Ile Gln Glu

225 230 235 240

Gly Ile Cys Gly Thr Thr Tyr Ser Asn Glu Val Ser Asp Ile Leu Lys
245 250 255

Val Asp Thr Ile Phe Gln Met Met Asn Thr Leu Lys Cys Pro Ser Leu
260 265 270

Lys Asp Lys Pro Lys Val Ile Ile Ile Gln Ala Cys Arg Gly Glu Lys
275 280 285

Gln Gly Val Val Leu Leu Lys Asp Ser Val Arg Asp Ser Glu Glu Asp
290 295 300

Phe Leu Thr Asp Ala Ile Phe Glu Asp Asp Gly Ile Lys Lys Ala His
305 310 315 320

Ile Glu Lys Asp Phe Ile Ala Phe Cys Ser Ser Thr Pro Asp Asn Val
325 330 335

Ser Trp Arg His Pro Val Arg Gly Ser Leu Phe Ile Glu Ser Leu Ile
340 345 350

Lys His Met Lys Glu Tyr Ala Trp Ser Cys Asp Leu Glu Asp Ile Phe
355 360 365

Arg Lys Val Arg Phe Ser Phe Glu Gln Pro Glu Phe Arg Leu Gln Met
370 375 380

Pro Thr Ala Asp Arg Val Thr Leu Thr Lys Arg Phe Tyr Leu Phe Pro
385 390 395 400

Gly His

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 840 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ACATTCTAAC TGCAACCTTT CGAAGCCTTT GCTCTGGCAC AACAGGTAGT AGGCGACACT
60

GTTCGTGTTG TCAACATGAC CAACAAGTGT CTCCTCCAAA TTGCTCTCCT GTTGTGCTTC
120

TCCACTACAG CTCTTTCCAT GAGCTACAAC TTGCTTGGAT TCCTACAAAG AAGCAGCAAT
180

TTTCAGTGTC AGAAGCTCCT GTGGCAATTG AATGGGAGGC TTGAATACTG CCTCAAGGAC
240

AGGATGAACT TTGACATCCC TGAGGAGATT AAGCAGCTGC AGCAGTTCCA GAAGGAGGAC
300

GCCGCATTGA CCATCTATGA GATGCTCCAG AACATCTTTG CTATTTTCAG ACAAGATTCA
360

TCTAGCACTG GCTGGAATGA GACTATTGTT GAGAACCTCC TGGCTAATGT CTATCATCAG
420

ATAAACCATC TGAAGACAGT CCTGGAAGAA AAACTGGAGA AAGAAGATTT CACCAGGGGA
480

AAACTCATGA GCAGTCTGCA CCTGAAAAGA TATTATGGGA GGATTCTGCA TTACCTGAAG
540

GCCAAGGAGT ACAGTCACTG TGCCTGGACC ATAGTCAGAG TGGAAATCCT AAGGAACTTT
600

TACTTCATTA ACAGACTTAC AGGTTACCTC CGAAACTGAA GATCTCCTAG CCTGTGCCTC
660

TGGGACTGGA CAATTGCTTC AAGCATTCTT CAACCAGCAG ATGCTGTTTA AGTGACTGAT
720

GGCTAATGTA CTGCATATGA AAGGACACTA GAAGATTTTG AAATTTTTAT TAAATTATGA
780

GTTATTTTTA TTTATTTAAA TTTTATTTTG GAAAATAAAT TATTTTTGGT GCAAAAGTCA
840

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 187 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
Met Thr Asn Lys Cys Leu Leu Gln Ile Ala Leu Leu Leu Cys Phe Ser
1               5                   10                  15

Thr Thr Ala Leu Ser Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg
            20                  25                  30

Ser Ser Asn Phe Gln Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg
        35                  40                  45

Leu Glu Tyr Cys Leu Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu
    50                  55                  60

Ile Lys Gln Leu Gln Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile
65                  70                  75                  80

Tyr Glu Met Leu Gln Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser
                85                  90                  95

Ser Thr Gly Trp Asn Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val
            100                 105                 110

Tyr His Gln Ile Asn His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu
        115                 120                 125

Lys Glu Asp Phe Thr Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys
    130                 135                 140

Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser
145                 150                 155                 160

His Cys Ala Trp Thr Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr
                165                 170                 175

Phe Ile Asn Arg Leu Thr Gly Tyr Leu Arg Asn
```

180 185

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1637 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GAGGCGAACC GGAGCGCGGG GCCGCGGTCG CCCCGACCAG AGCCGGGAGA CCGCAGCACC 60

CGCAGCCGCC CGCGAGCGCG CCGAAGACAG CGCGCAGGCG AGAGCGCGCG GGCGGGGGCG 120

CGCAGGCCCT GCCCGCCCCT TCCGTCCCCA CCCCCCTCCG CCCTTTCCTC TCCCCACCTT 180

CCTCTCGCCT CCCGCGCCCC CGCACCGGGC GCCCACCCTG TCCTCCTCCT GCGGGAGCGT 240

TGTCCGTGTT GGCGGCCGCA GCGGGCCGGG CCGGTCCGGC GGGCCGGGGG ATGGCGCTGC 300

TGGACCTGGC CTTGGAGGGA ATGGCCGTCT TCGGGTTCGT CCTCTTCTTG GTGCTGTGGC
360

TGATGCATTT CATGGCTATC ATCTACACCC GATTACACCT CAACAAGAAG GCAACTGACA
420

AACAGCCTTA TAGCAAGCTC CCAGGTGTCT CTCTTCTGAA ACCACTGAAA GGGGTAGATC
480

CTAACTTAAT CAACAACCTG GAAACATTCT TTGAATTGGA TTATCCCAAA TATGAAGTGC
540

TCCTTTGTGT ACAAGATCAT GATGATCCAG CCATTGATGT ATGTAAGAAG CTTCTTGGAA
600

AATATCCAAA TGTTGATGCT AGATTGTTTA TAGGTGGTAA AAAAGTTGGC ATTAATCCTA
660

AAATTAATAA TTTAATGCCA GGATATGAAG TTGCAAAGTA TGATCTTATA TGGATTTGTG
720

ATAGTGGAAT AAGAGTAATT CCAGATACGC TTACTGACAT GGTGAATCAA ATGACAGAAA
780

AAGTAGGCTT GGTTCACGGG CTGCCTTACG TAGCAGACAG ACAGGGCTTT GCTGCCACCT
840

TAGAGCAGGT ATATTTTGGA ACTTCACATC CAAGATACTA TATCTCTGCC AATGTAACTG
900

GTTTCAAATG TGTGACAGGA ATGTCTTGTT TAATGAGAAA AGATGTGTTG GATCAAGCAG
960

GAGGACTTAT AGCTTTTGCT CAGTACATTG CCGAAGATTA CTTTATGGCC AAAGCGATAG
1020

CTGACCGAGG TTGGAGGTTT GCAATGTCCA CTCAAGTTGC AATGCAAAAC TCTGGCTCAT
1080

ATTCAATTTC TCAGTTTCAA TCCAGAATGA TCAGGTGGAC CAAACTACGA ATTAACATGC
1140

TTCCTGCTAC AATAATTTGT GAGCCAATTT CAGAATGCTT TGTTGCCAGT TTAATTATTG
1200

GATGGGCAGC CCACCATGTG TTCAGATGGG ATATTATGGT ATTTTTCATG TGTCATTGCC
1260

TGGCATGGTT TATATTTGAC TACATTCAAC TCAGGGGTGT CCAGGGTGGC ACACTGTGTT
1320

TTTCAAAACT TGATTATGCA GTCGCCTGGT TCATCCGCGA ATCCATGACA ATATACATTT
1380

TTTTGTCTGC ATTATGGGAC CCAACTATAA GCTGGAGAAC TGGTCGCTAC AGATTACGCT
1440

GTGGGGGTAC AGCAGAGGAA ATCCTAGATG TATAACTACA GCTTTGTGAC TGTATATAAA
1500

GGAAAAAAGA GAAGTATTAT AAATTATGTT TATATAAATG CTTTTAAAAA TCTACCTTCT
1560

GTAGTTTTAT CACATGTATG TTTTGGTATC TGTTCTTTAA TTTATTTTTG CATGGCACTT
1620

GCATCTGTGA AAAAAAA
1637

---

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 394 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Ala Leu Leu Asp Leu Ala Leu Glu Gly Met Ala Val Phe Gly Phe
1 5 10 15

Val Leu Phe Leu Val Leu Trp Leu Met His Phe Met Ala Ile Ile Tyr
20 25 30

Thr Arg Leu His Leu Asn Lys Lys Ala Thr Asp Lys Gln Pro Tyr Ser
35 40 45

Lys Leu Pro Gly Val Ser Leu Leu Lys Pro Leu Lys Gly Val Asp Pro
50 55 60

Asn Leu Ile Asn Asn Leu Glu Thr Phe Phe Glu Leu Asp Tyr Pro Lys
65 70 75 80

Tyr Glu Val Leu Leu Cys Val Gln Asp His Asp Asp Pro Ala Ile Asp
85 90 95

Val Cys Lys Lys Leu Leu Gly Lys Tyr Pro Asn Val Asp Ala Arg Leu
100 105 110

Phe Ile Gly Gly Lys Lys Val Gly Ile Asn Pro Lys Ile Asn Asn Leu

115 120 125

Met Pro Gly Tyr Glu Val Ala Lys Tyr Asp Leu Ile Trp Ile Cys Asp
130 135 140

Ser Gly Ile Arg Val Ile Pro Asp Thr Leu Thr Asp Met Val Asn Gln
145 150 155 160

Met Thr Glu Lys Val Gly Leu Val His Gly Leu Pro Tyr Val Ala Asp
165 170 175

Arg Gln Gly Phe Ala Ala Thr Leu Glu Gln Val Tyr Phe Gly Thr Ser
180 185 190

His Pro Arg Tyr Tyr Ile Ser Ala Asn Val Thr Gly Phe Lys Cys Val
195 200 205

Thr Gly Met Ser Cys Leu Met Arg Lys Asp Val Leu Asp Gln Ala Gly
210 215 220

Gly Leu Ile Ala Phe Ala Gln Tyr Ile Ala Glu Asp Tyr Phe Met Ala
225 230 235 240

Lys Ala Ile Ala Asp Arg Gly Trp Arg Phe Ala Met Ser Thr Gln Val
245 250 255

Ala Met Gln Asn Ser Gly Ser Tyr Ser Ile Ser Gln Phe Gln Ser Arg
260 265 270

Met Ile Arg Trp Thr Lys Leu Arg Ile Asn Met Leu Pro Ala Thr Ile
275 280 285

Ile Cys Glu Pro Ile Ser Glu Cys Phe Val Ala Ser Leu Ile Ile Gly
290 295 300

Trp Ala Ala His His Val Phe Arg Trp Asp Ile Met Val Phe Phe Met
305 310 315 320

Cys His Cys Leu Ala Trp Phe Ile Phe Asp Tyr Ile Gln Leu Arg Gly
325 330 335

Val Gln Gly Gly Thr Leu Cys Phe Ser Lys Leu Asp Tyr Ala Val Ala
340 345 350

Trp Phe Ile Arg Glu Ser Met Thr Ile Tyr Ile Phe Leu Ser Ala Leu
355 360 365

Trp Asp Pro Thr Ile Ser Trp Arg Thr Gly Arg Tyr Arg Leu Arg Cys
370 375 380

Gly Gly Thr Ala Glu Glu Ile Leu Asp Val
385 390

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2599 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CAGATTCACA AACTGCAGGA CTGGGCAGGG AGCAGACAGT GAGCAAACGC CAGCAGGGCT
60

GCTGTGAATT TGTGTAAGGA TTGAGGGACA GTTGCTTTTC AGCATGGGCC CAGGAATGCC
120

AAGGAGACAT CTATGCACGA CCTTGGGAAA TGAGTTGATG TCTCCGGTAA AACACCGGAG
180

ACTAATTCCT GCCCTGCCCA ATTTTGCAGG GAGCATGGCT GTGAGGATGG GGTGAACTCA
240

CGCACAGCCA AGGACTCCAA AATCACAACA GCATTACTGT TCTTATTTGC TGCCACACCT
300

GAGCCAGCCT GCTCCTTCCC AGGAGTGGAG GAGGCCTGGG GGGAGGGAGA GGAGTGACTG
360

AGCTTCCCTC CCGTGTGTTC TCCGTCCCTG CCCCAGCAAG ACAACTTAGA TCTCCAGGAG
420

AACTGCCATC CAGCTTTGGT GCAATGGCTG AGTGCACAAG TGAGTTGTTG CCCTGGGTTT
480

CTTTAATCTA TTCAGCTAGA ACTTTGAAGG ACAATTTCTT GCATTAATAA AGGTTAAGCC
540

CTGAGGGGTC CCTGATAACA ACCTGGAGAC CAGGATTTTA TGGCTCCCCT CACTGATGGA
600

CAAGGAGGTC TGTGCCAAAG AAGAATCCAA TAAGCACATA TTGAGCACTT GCTGTATATG
660

CAGTATTGAG CACTGTAGGC AAGACCCAAG AAAGAGAAGG AGCCATCTCC ATCTTGAAGG
720

AACTCAAAGA CTCAAGTGGG AACGACTGGG CACTGCCACC ACCAGAAAGC TGTTCGACGA
780

GACGGTCGAG CAGGGTGCTG TGGGTGATAT GGACAGCAGA AGGGGGAGAC CAAGGTTCCA 840

GCTCAACCAA TAACTATTGC ACAACCACCT GTCCCTGCCT CAGTTCCCTT TTATGTAACA 900

TGAAGTCGTT GTGAGGGTTA AAGGCAGTAA CAGGTATAAA GTACTTAGAA AAGCAAAGGG 960

TGCTACGTAC ATGTGAGGCA TCATTACGCA GACGTAACTG GGATATGTTT ACTATAAGGA 1020

AAAGACACTG AGGTCTAGAA ATAGCTCCGT GGAGCAGAAT CAGTATTGGG AGCCGGTGGC 1080

GGTGTGAAGC ACCAGTGTCT GGCACACAGT AGGTGCTCAT TGGCTCCCTT CCACCTGTCA 1140

TTCCCACCAC CCTGAGGCCC CAACCGCCAC ACACACAGGA GCATTTGGAG AGAAGGCCAT 1200

GTCTTCAAAG TCTGATTTGT GATGAGGCAG AGGAAGATAT TTCTAATCGG TCTTGCCCAG 1260

AGGATCACAG TGCTGAGACC CCCCACCACC AGCCGGTACC TGGGAAGGGG GAGAGTGCAG 1320

GCCTGCTCAG GGACTGTTCC TGTCTCAGCA ACCAAGGGAT TGTTCCTGTC AATCAATGGT 1380

TTATTGGAAG GTGGCCCAGT ATGAGCCCTA GAAGAGTGTG AAAAGGAATG GCAATGGTGT 1440

TCACCATCGG CAGTGCCAGG GCAGCACTCA TTCACTTGAT AAATGAATAT TTATTAGCTG 1500

GTTGGAGAGC TAGAACCTGG AGAGCTAGAA CCTGGAGAAC TAGAACCTGG AGGGCTAGAA
1560

CCTGGAGAGG CTAGAACCAA GAAGGGCTAG AACCTGGAGG GGCTAGAACC TAGAGAAGCT
1620

AAAACCTGAG CTAGAAGCTG GAGGACTAGA ACCTGGAGGG CTGGAATCTG AAGGGCTAGA
1680

ACCTGGAGGG CTGGAATCTG GAGAGCTAGA ACCTGGAGGG CTAGAACCTG GAGGGCTAGA
1740

ACCTAGAAGG GCTAGAACCT GGAGGGCTGG AATCTGGAGA GCTAGAACCT GGAGGGCTAG
1800

AACCTGGAGG GCTAGAACCT AGAAGGGCTA GAACCTGGAG GGCTAGAACC TGGCAGGTTA
1860

GAACCTAGAA GGGCTAGAAC CTGGAGAGCC AGAACCTGGA GGGCTAGAAC CTGGAAGGGC
1920

TAGAACCTGT AGAGCTAGAA CATGGAGAGC TAGAACCCGG CAGGCTAGAA CCTGGCAAGC
1980

TAGAACCTGG AGGGAATGAA CCTGGAGGGC TAGAACCTGG AGAATGAGAA AAATTTACAT
2040

GGCAAAGAGC CCATAAATCC TGACCAATCC AACTCTGAAT TTTAAAGCAA AAGCGTGAAA
2100

AAAAAGATTC CCTCCTTACC CCCAACCCAC TCTTTTTTCC CACCACCCAC TCTCCTCTGC
2160

CTCAGTAAGT ATCTGGAGGA AGAAAACAGG TGAAAGAAGA AGTAAAAACC ATTTAGTATT
2220

AGTATTAGAA TGAAGTCAAA CTGTGCCACA CATGGTGAAT GAAAAAAAAA AAAAAGAGGC
2280

TGTGTTTTGT CACACAGGGC AGTCATTCAG CACCAGAGCA CGTGATGGTC TGAGACTCTC
2340

TTAGGAGCAG AGCTCTGCCG CAATGGCCAT GTGGGGATCC ACACCTGGTC TGAGGGGCAA
2400

CTGAGTCTGC GGGAGAAGAG CGGCCCTATG CATGGTGTAG ATGCCCTGAT AAAGAACATC
2460

TGTCCTGTGA AAGACTCAAT GAGCTGTTAT GTTGTAAACA GGAAGCATTT CACATCCAAA
2520

CGAGAAAATC ATGTAAACAT GTGTCTTTTC TGTAGAGCAT AATAAATGGA TGAGGTTTTT
2580

GCAAAAAAAA AAAAAAAAA
2599

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: C-terminal

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Gln Ile His Lys Leu Gln Asp Trp Ala Gly Ser Arg Gln
1 5 10

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1072 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GAATTCGGCA CGAGGCTATT ACAAGTTTAG AAAAAACAAA GCAATTGTCA AAAAAAGTTA
60

GAACTATTAC AACCCCTGTT TCCTGGTACT TATCAAATAC TTAGTATCAT GGGGGTTGGG
120

AAATGAAAAG TAGGAGAAAA GTGAGATTTT ACTAAGACCT GTTTTACTTT ACCTCACTAA
180

CAATGGGGGG AGAAAGGAGT ACAAATAGGA TCTTTGACCA GCACTGTTTA TGGCTGCTGT
240

GGTTTCAGAG AATGTTTATA CATTATTTCT ACCGAGAATT AAAACTTCAG ATTGTTCAAC
300

ATGAGAGAAA GGCTCAGCAA CGTGAAATAA CGCAAATGGC TTCCTCTTTC CTTTTTTGGA
360

CCATCTCAGT CTTTATTTGT GTAATTCATT TTGAGGAAAA AACAACTCCA TGTATTTATT
420

CAAGTGCATT AAAGTCTACA ATGGAAAAAA AGCAGTGAAG CATTAGATGC TGGTAAAAGC
480

TAGAGGAGAC ACAATGAGCT TAGTACCTCC AACTTCCTTT CTTTCCTACC ATGTAACCCT
540

GCTTTGGGAA TATGGATGTA AAGAAGTAAC TTGTGTCTCC ATGGAAAATC AGTACCAATC
600

ACACCAAGGG AGGATGAAAC CGCCGGAACA AAAATGAGGT GTGTAGAACA GGGTCCCACA
660

GGTTTGGGGA CATTGAGATC ACTTGTCTTG TGGTGGGGAG GCTGCTGAGG GGTAGCAGGT
720

CCATCTCCAG CAGCTGGTCC AACAGTCGTA TCCTGGTGAA TGTCTGTTCA GCTCTTCTGT
780

GAGAATATGA TTTTTTCCAT ATGTATATAG TAAAATATGT TACTATAAAT TACATGTACT
840

TTATAAGTAT TGGTTTGGGT GTTCCTTCCA AGAAGGACTA TAGTTAGTAA TAAATGCCTA
900

TAATAACATA TTTATTTTTA TACATTTATT TCTAATGAAA AAAACTTTTA AATTATATCG
960

CTTTTGTGGA AGTGCATATA AAATAGAGTA TTTATACAAT ATATGTTACT AGAAATAAAA 1020

GAACACTTTT GGAAAAAAAA AAAAAAAAAA AAATTGCCGC CGCAAGCTTA AT 1072

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 672 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GGACACAATA GAATGAGCCA ACATGATGGT TTCTCTCCAG TAAGAGTTTT TCTTTTGGAA 60

ATGAGGTTAA CCTAGCCCCA AATCTAGCAA TTCTCATAAA ATCCGATTTT AGAATTAGCC 120

TCCCAGATTA ATCTGAATGA TTGACTTATT TTTTCTTAGG CAAGTCAGTA AGCCACCCAC 180

TAGACAGCCA TATCCAGCAA AATAAGAGAA GTTTCCAGAT GCCAAATGAT AAGCCACCAT 240

CAACCCAGCG GGGAAGCCTT CTGGTTGGTT TGGCTGTATG AGATTCAGGA AGGCCAGAAT
300

ACCCAAAATT ATTCACACGA CGTTAACTTA TTGGTACTGG CTAAGCAATA CATGTATTTC
360

CTAAAGGAGG AGATGGTCTT TTGGTTGATT TATGGACACA CTTGTTTCAT CTGACTGTAA
420

ATATATTGCA TGCTTTATTC TGATGGTGCA CTATTTCATC CAGCAAGCTT TTCATCTGAG
480

AATGTTTAAT GTTGACCTTA TTCTTAGAGC AAGTAGATCT AAATATTTTT CAGCTGAGTT
540

ATTAGGGAGT CATTATTCTG TGGTACAATG CTGCAAAAAG CATCATGTGG AAGAATGGGA
600

ACTATGCTTA CTTTATGAAG TGATGTATAA CACAATGAAA TCTGTTTTAC AACTACAAAA
660

AAAAAAAAAA AA
672